OM of: US-09-303-518D-653 to: A_Geneseq_032802:* out_format : pfs

Date: Jun 30, 2002 7:14 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09303518/runat_28062002_142712_4291/app_query.fasta_1.23501
-DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09303518_@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518D-653
Query length: 4407
Database: A_Geneseq_032802:*
Database sequences: 747574
Database length: 111073796
Search time (sec): 627.340000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation	..
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:AAR07304	+	1614.00	2049.86	3.2e-107	15		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75564	+	1585.50	2013.94	3.4e-105	14		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB52592	+	1052.50	1327.60	6.1e-67	13		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB52745	+	1048.00	1321.84	1.3e-66	13		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW27705	+	721.50	916.78	2.0e-43	32		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:AAP80136	+	442.50	548.31	2.9e-23	74		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG98842	+	359.00	432.61	3.8e-17	15		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01830	+	310.50	372.82	1.0e-13	12		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01828	+	310.50	372.77	1.0e-13	12		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41731	+	308.50	369.26	1.5e-13	13		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY81609	+	306.00	366.89	2.2e-13	12		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30291	+	306.00	364.12	2.4e-13	15		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41732	+	296.50	352.36	1.1e-12	15		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30292	+	296.50	351.86	1.2e-12	16		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41728	+	287.00	340.49	5.4e-12	14		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR63506	+	286.50	339.85	5.9e-12	14		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01825	+	283.50	338.04	9.0e-12	12		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01824	+	283.00	337.34	9.8e-12	12		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01848	+	282.50	334.69	1.1e-11	14		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:AAR26503	+	279.00	335.41	1.7e-11	91		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41724	+	278.50	329.54	2.2e-11	14		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:AAR25578	+	275.50	330.77	3.0e-11	92		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:AAR14320	+	274.00	328.97	3.9e-11	91		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30294	+	273.50	323.10	5.1e-11	14		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01827	+	273.00	327.01	4.7e-11	96		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01826	+	273.00	326.94	4.7e-11	97		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01841	+	272.00	325.34	5.6e-11	10		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01840	+	272.00	325.27	5.6e-11	10		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75097	+	271.00	314.12	9.4e-11	25		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01837	+	269.50	321.40	8.6e-11	10		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01836	+	269.50	321.34	8.7e-11	10		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR63505	+	268.50	316.23	1.2e-10	15		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01846	+	268.50	316.23	1.2e-10	15		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37017	+	267.50	300.58	2.3e-10	57		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30293	+	267.00	314.30	1.5e-10	15		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01845	+	266.00	315.86	1.6e-10	11		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01844	+	266.00	315.79	1.6e-10	11		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41725	+	266.00	313.01	1.8e-10	15		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37120	+	266.00	308.44	2.1e-10	23		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01835	+	265.50	316.02	1.7e-10	10		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01834	+	265.50	315.97	1.7e-10			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41723	+	265.00	311.72	2.1e-10			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU34143	+	261.00	303.26	4.6e-10			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37018	+	261.00	298.77	5.4e-10			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37403	+	260.50	290.69	7.6e-10			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG20365	+	258.00	306.33	5.8e-10			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01833	+	257.50	306.64	6.1e-10			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01832	+	257.50	306.58	6.1e-10			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB23860	+	257.50	297.18	8.6e-10			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01847	+	256.50	304.43	7.5e-10			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01843	+	256.00	304.85	7.8e-10			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01842	+	256.00	304.79	7.8e-10			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75096	+	254.00	294.44	1.4e-09			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAR99393	+	253.50	292.29	1.6e-09			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826	+	253.00	295.72	1.5e-09			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU27577	+	251.50	294.97	1.9e-09			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAE10024	+	251.50	294.97	1.9e-09			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB52597	+	248.00	297.20	2.7e-09			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB69135	+	247.50	286.03	4.2e-09			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01849	+	244.00	288.92	5.8e-09			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW65088	+	243.50	283.50	7.4e-09			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY27230	+	243.50	281.28	8.0e-09			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU34320	+	237.50	271.11	2.4e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37374	+	237.50	271.11	2.4e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01839	+	237.00	280.76	1.8e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01838	+	237.00	280.69	1.8e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB61077	+	237.00	278.32	1.9e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU34139	+	236.50	270.32	2.7e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU36796	+	236.50	270.32	2.7e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG98256	+	235.00	274.66	2.8e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG30355	+	233.00	270.01	4.2e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW42634	+	232.50	267.09	5.0e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU34339	+	230.00	261.64	8.1e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75098	+	230.00	260.93	8.3e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB69136	+	229.00	260.90	9.4e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU27578	+	228.50	265.75	8.4e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAW04505	+	228.50	261.88	9.6e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB69133	+	228.50	261.88	9.6e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB69137	+	228.50	261.88	9.6e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB69134	+	228.50	261.58	9.7e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAE00701	+	228.50	261.19	9.9e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU34338	+	228.00	259.05	1.1e-07			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAE10025	+	227.50	264.47	9.9e-08			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59813	+	226.50	266.57	1.0e-07			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY95550	+	224.00	265.03	1.5e-07			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU38896	+	224.00	258.39	1.9e-07			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:AAR05041	+	224.00	249.54	2.6e-07			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:AAR04232	+	223.50	260.33	1.8e-07			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW56322	+	223.00	254.54	2.4e-07			

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:AAR07304

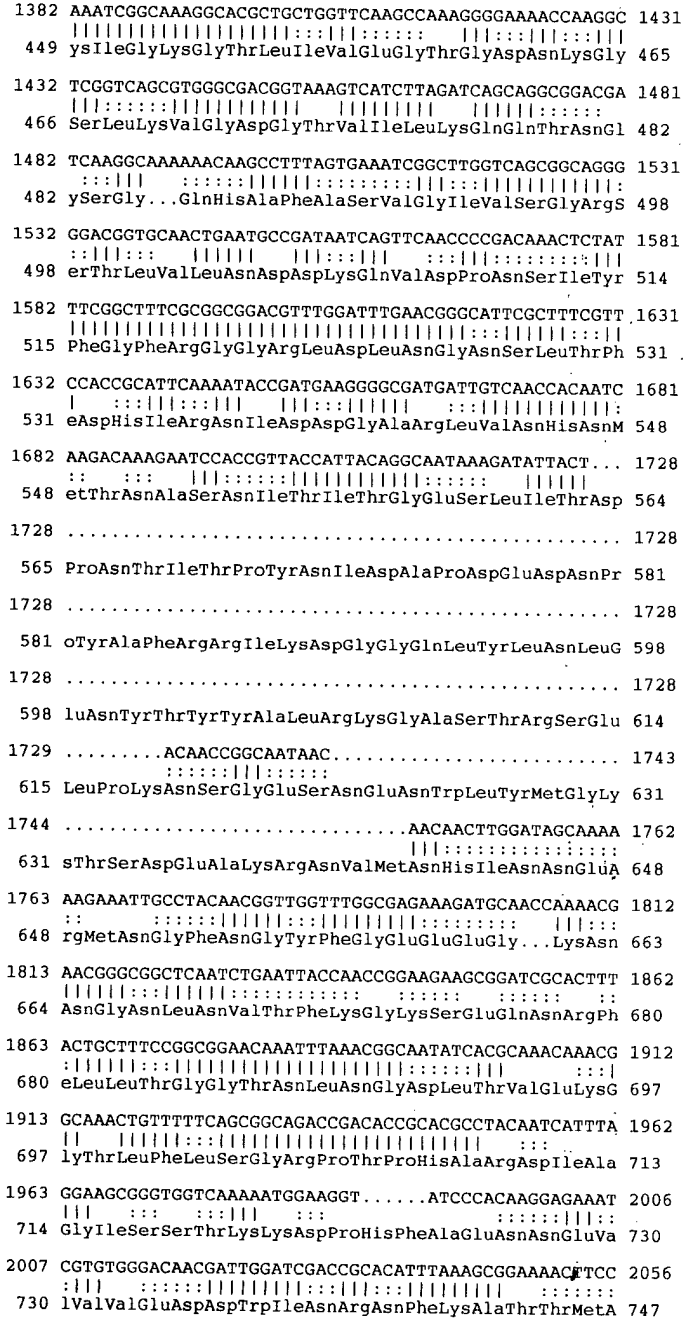
seq_documentation_block:
ID AAR07304 standard; protein; 1541 AA.
XX
AC AAR07304;
XX
DT 31-JAN-1991 (first entry)
XX
DE IgA1 protease.
XX
KW IgA1; vaccine; meningitis; gonnorhoea; allergies.
XX
OS Haemophilus influenzae.
XX
PN WO9011367-A.
XX
PD 04-OCT-1990.
XX
PF 16-MAR-1990; 90WO-DK00073.
XX
PR 17-MAR-1989; 89DK-0001308.
XX

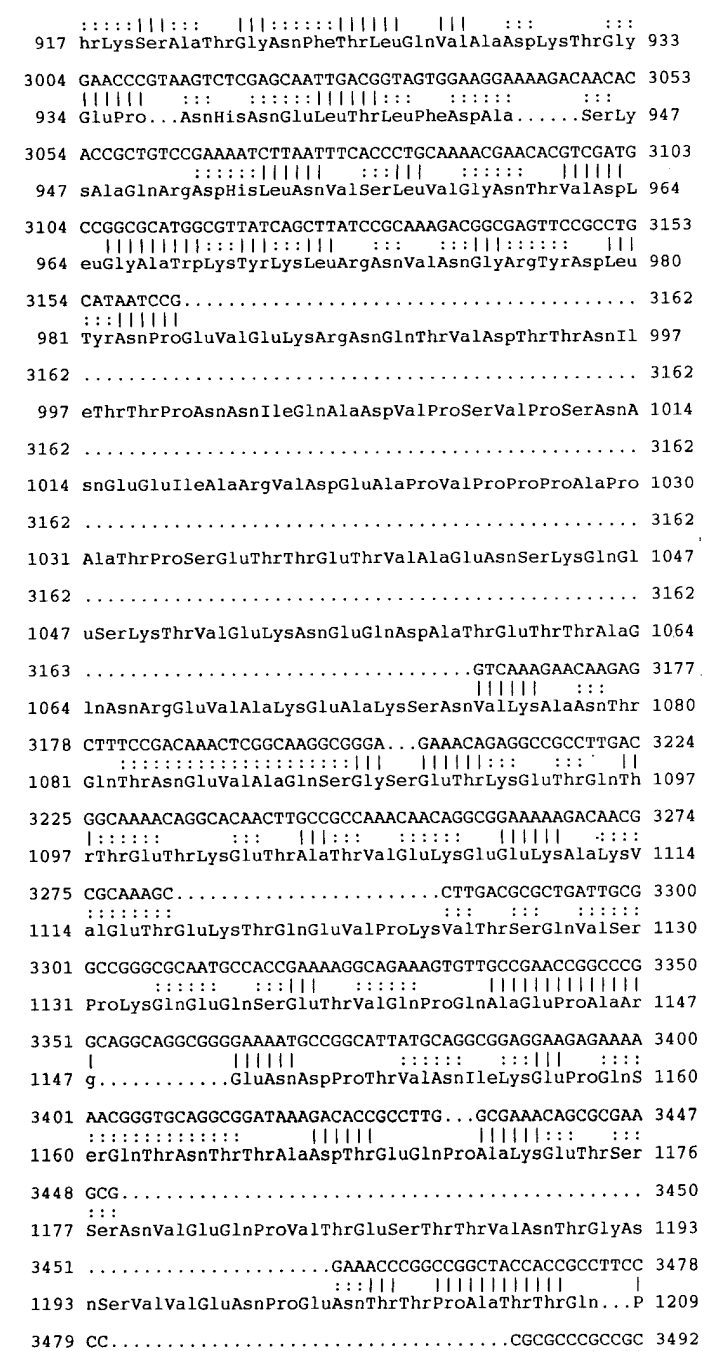
```
PA   (KILI/) KILIAN M.
XX
PI   Kilian M,  Poulsen K;
XX
DR   WPI; 1990-320267/42.
DR   N-PSDB; AAQ06164.
XX
PT   Immunoglobulin A1 protease prodn. - by cloning from
PT   microorganisms for immunisation against immunoglobulin A1
PT   protease producing bacteria
XX
PS   Disclosure; fig 3; 44pp; English.
XX
CC   This immunoglobulin (Ig)A1 protease is produced by recombinant DNA
CC   methods.  It is useful in a vaccine for e.g. meningococcal meningitis,
CC   gonorrhoea or allergic diseases.  It specifically cleaves the heavy
CC   chain of human IgA1 in the hinge region.
XX
SQ   Sequence   1541 AA;
```

```
alignment_scores:
            Quality: 1614.00              Length:   1716
              Ratio:  1.732                Gaps:     48
 Percent Similarity:  54.312   Percent Identity: 26.865

alignment_block:
 US-09-303-518D-653 x AAR07304   ..

 Align seg 1/1  to: AAR07304  from: 1  to: 1541
```

```
     64 CGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCC 113
        :::|||:::      :::::::|||:::   ::::::::::::::   ||
      5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

    114 CCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
        |   ::::::|||:::   :::         ::::::||||||   :::|
     21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

    164 GCGACTTTGCCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
        ||||||||||||||||||||||||||:::||||||||||   ::::::
     38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54

    214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGC 263
           |||   :::|||:::::::::   :::|||   ::::::   :::::
     55 LeuValLysAspLysAsnAsnLysAspLeuGlyThrAlaLeuProAsnGl 71

    264 C...CCGATGATTGATTTTTCTGTGGTATCG...CGTAACGGCGTGGCGG 307
        :   |||||||||||||||||||||||:::      :::   :::|||:
     71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

    308 CATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT...AACGGCGGC 354
        ::|||   :::   ||||||:::|||:::|||   |||   :::::::|||
     88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

    355 TATAACAATGTTGATTTTGGT..............GCGGAGGGAAGCAA 389
           :::::::::   ||||||                :::::::::||
    105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

    390 TCCCGATCAGCACCGC.................TTTTCTTACCAAATTG 421
        |         ||||||                        |||      |
    121 nAlaLysAlaHisArgAspValSerSerGluGluAsnArgTyrPheSerV 138

    422 TGAAAAGAAATAATTATAAAGCAGGGACTAACGGC.............. 456
        ||:::::::|||:::|||   :::      ||||||
    138 alGluLysAsnGluTyrProThrLysLeuAsnGlyLysThrValThrThr 154

    457 ............CATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
                     :::            ||||||:::||||||||||||
    155 GluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171

    495 CAAATTTGTCACAGATGCAGAACCTGTTGAG...ATGACCAGTTATATGG 541
         ||||||||||||:::::   |||:::|||      |||:::      |
```

```
    171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaSerSerA 188

    542 ATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC 591
        ||:::      |||   |||   ||||||||||||      ||||||:::
    188 spAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204

    592 GGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAA........... 630
        |||:::|||   |||:::      :::      |||:::
    205 GlySerGlySerGlnPheIleTyrLysLysGlyAspAsnTyrSerLeuIl 221

    631 .CCCAATAACCGCGAAAGTTCATATCAT............ATTGCAAGCG 667
           ||||||:::|||   :::   :::            :::::::::|
    221 eLeuAsnAsnHisGluValGlyGlyAsnAsnLeuLysLeuValGlyAspA 238

    668 CATATTCTTGGCTCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGT 717
        |||||::::::   ::::::|||:::   :::      |||   ::::::
    238 laTyrThrTyrGlyIleAlaGlyThrProTyrLysValAsnHisGluAsn 254

    718 GGTGGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGC...CCATA 764
        :::|||   :::::::::|||::::::|||   :::|||||   |||
    255 AsnGlyLeuIleGlyPheGlyAsnSerLysGluGluHisSerAspProLy 271

    765 TGGTTTTTACCAACAGGA..............GGCTCATTTGGCGACA 799
         |||:::|||                      :::   :::|||||||
    271 sGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyAspS 288

    800 GTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATT 849
        ||||||||||:::|||:::|||||   :::|||   |||||||||:::
    288 erGlySerProLeuPheValTyrAspArgGluLysGlyLysTrpLeuPhe 304

    850 AATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTT 899
           |||      :::              |||      ||||||
    305 LeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLysSer........ 318

    900 CCAGCTAGTTCGTAAAGATTGG...TTCTATGATGAAATCTTTGCTGGAG 946
                     :::::::|||   :::|||   :::   |||:::   |
    319 ..........TrpGlnGluTrpAsnIleTyrLysSerGlnPheThrLysA 332

    947 ATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTT 996
        ||:::
    332 spValLeu..................................... 334

    997 AACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACACTA 1046
              |||:::::::::|||||:::::   ::::::          ||
    335 ......AsnLysAspSerAlaGlySerLeuIleGlySerLysThrAspTy 349

   1047 TTCTCTACCTTATAGATTAAAAACACGAACCGTT............CAAT 1084
        ||||          :::   ||||||   |||:::            :::
    349 rSerTrpSerSerAsnGlyLysThrSerThrIleThrGlyGlyGluLysS 366

   1085 TGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCT 1134
          :::||||||:::|||::::::          :::|||
    366 erLeuAsnValAspLeuAlaAspGlyLysAspLysPro............ 378

   1135 GCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATAT 1184
                                          |||:::|||::::::::
    379 ...............................AsnHisGlyLysSerVa 384

   1185 TTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATCA 1234
        ::::|||      :::|||:::|||   |||   |||::::::||||||:
    384 lThrPhe......GluGlySerGlyThrLeuThrLeuAsnAsnAsnIleA 399

   1235 ACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTC...TCG 1281
        ::|||||||||||||||||:::||||||||::::::   |||   :::
    399 spGlnGlyAlaGlyGlyLeuPhePheGluGlyAspTyrGluValLysGly 415

   1282 CCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGG 1331
          :::::::|||   ||||||::||||||||||||   :::::::::||
    416 ThrSerAspAsnThrThrTrpLysGlyAlaGlyValSerValAlaGluGl 432

   1332 CAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCA 1381
        |:::|||||||||||||||||::::::          |||||||||:::|
    432 yLysThrValThrTrpLysValHisAsnProGlnTyrAspArgLeuAlaL 449
```

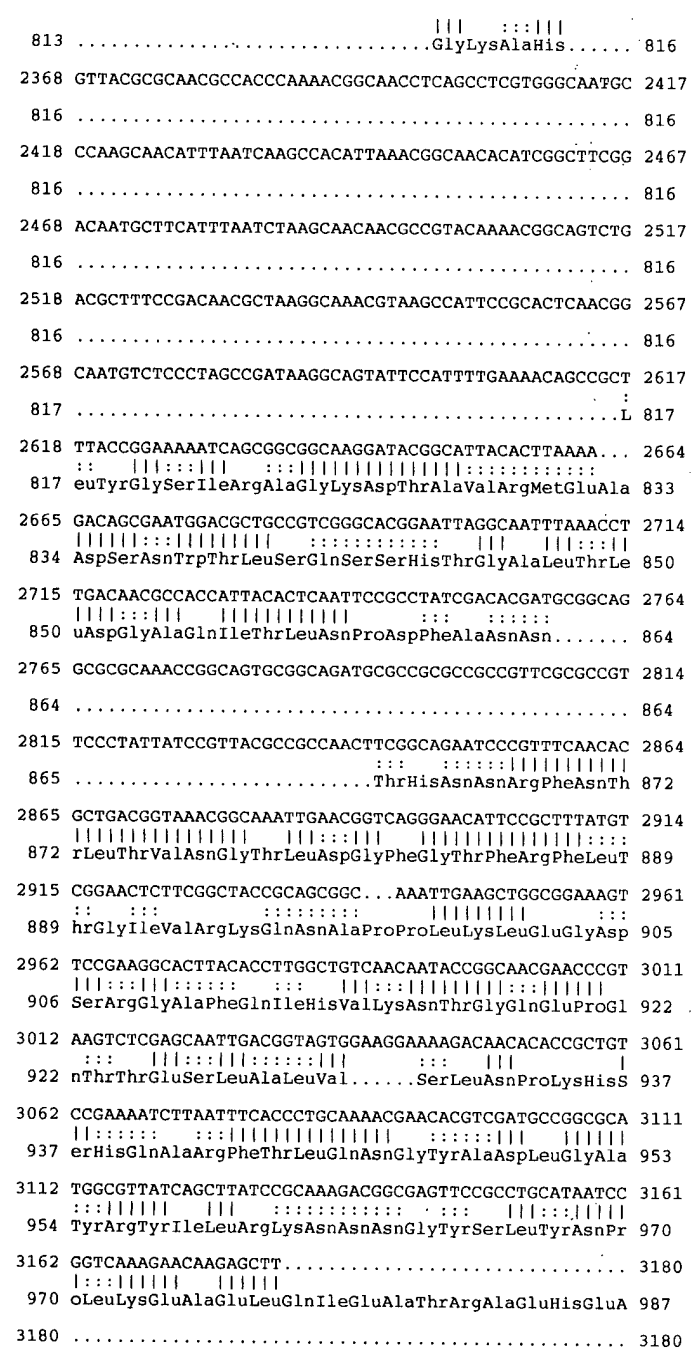

1382 AAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGC 1431
449 ysIleGlyLysGlyThrLeuIleValGluGlyThrGlyAspAsnLysGly 465
1432 TCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGA 1481
466 SerLeuLysValGlyAspGlyThrValIleLeuLysGlnGlnThrAsnGl 482
1482 TCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGG 1531
482 ySerGly...GlnHisAlaPheAlaSerValGlyIleValSerGlyArgS 498
1532 GGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTAT 1581
498 erThrLeuValLeuAsnAspAspLysGlnValAspProAsnSerIleTyr 514
1582 TTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTT 1631
515 PheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnSerLeuThrPh 531
1632 CCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATC 1681
531 eAspHisIleArgAsnIleAspAspGlyAlaArgLeuValAsnHisAsnM 548
1682 AAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACT... 1728
548 etThrAsnAlaSerAsnIleThrIleThrGlyGluSerLeuIleThrAsp 564
1728 1728
565 ProAsnThrIleThrProTyrAsnIleAspAlaProAspGluAspAsnPr 581
1728 1728
581 oTyrAlaPheArgArgIleLysAspGlyGlyGlnLeuTyrLeuAsnLeuG 598
1728 1728
598 luAsnTyrThrTyrTyrAlaLeuArgLysGlyAlaSerThrArgSerGlu 614
1729 .....ACAACCGGCAATAAC..... 1743
615 LeuProLysAsnSerGlyGluSerAsnGluAsnTrpLeuTyrMetGlyLy 631
1744 .....AACAACTTGGATAGCAAAA 1762
631 sThrSerAspGluAlaLysArgAsnValMetAsnHisIleAsnAsnGluA 648
1763 AAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACG 1812
648 rgMetAsnGlyPheAsnGlyTyrPheGlyGluGluGluGly...LysAsn 663
1813 AACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTT 1862
664 AsnGlyAsnLeuAsnValThrPheLysGlyLysSerGluGlnAsnArgPh 680
1863 ACTGCTTTCCGGCGGAACAAATTTAAACGGCAATATCACGCAAACAAACG 1912
680 eLeuLeuThrGlyGlyThrAsnLeuAsnGlyAspLeuThrValGluLysG 697
1913 GCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTA 1962
697 lyThrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIleAla 713
1963 GGAAGCGGGTGGTCAAAAATGGAAGGT......ATCCCACAAGGAGAAAT 2006
714 GlyIleSerSerThrLysLysAspProHisPheAlaGluAsnAsnGluVa 730
2007 CGTGTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCC 2056
730 lValValGluAspAspTrpIleAsnArgAsnPheLysAlaThrThrMetA 747
2057 ATATTCAGGGCGGACAAGCGGTGGTTTCC...CGCAATGTTGCCAAAGTG 2103
747 snValThrGlyAsnAlaSerLeuTyrSerGlyArgAsnValAlaAsnIle 763
2104 GAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGC 2153
764 ThrSerAsnIleThrAlaSerAsnLysAlaGlnValHisIleGlyTyr.. 779
2154 ACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGA 2203
780 ....LysThrGlyAspThrValCysValArgSerAspTyrThrGlyTyrV 795
2204 CAAGTTGTACCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTG 2253
795 alThrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPhe 810
2254 AGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTT 2303
811 AsnProThrAsnLeuArgGlyAsnValAsnLeuThrGluSerAlaAsnPh 827
2304 AAATCTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGCAGGCGGAG 2353
827 eValLeu...GlyLysAlaAsnLeuPheGlyThrIleGlnSerArgGlyA 843
2354 ACACGCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACCTCAGC 2403
843 snSerGlnValArgLeuThrGluAsn..... 851
2404 CTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAA 2453
851 851
2454 CACATCGGCTTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTAC 2503
851 851
2504 AAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCAT 2553
851 851
2554 TCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTT 2603
851 851
2604 TGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCAT 2653
851 851
2654 TACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGC 2703
852 .....SerHisTrpHisLeuThrGlyAsnSerAspValHis 863
2704 AATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACA 2753
864 GlnLeuAspLeuAlaAsnGlyHisIleHisLeuAsnSerAlaAspAsnSe 880
2754 CGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCC 2803
880 rAsn..... 881
2804 GTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCC 2853
882 .....AsnValThr..... 884
2854 CGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATT 2903
885 LysTyrAsnThrLeuThrValAsn...SerLeuSerGlyAsnGlySerPh 900
2904 CCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGG 2953
900 eTyrTyrLeuThrAspLeuSerAsnLysGlnGlyAspLysValValValT 917
2954 CGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAAC 3003


917 hrLysSerAlaThrGlyAsnPheThrLeuGlnValAlaAspLysThrGly 933
3004 GAACCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACAC 3053
934 GluPro...AsnHisAsnGluLeuThrLeuPheAspAla.....SerLy 947
3054 ACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATG 3103
947 sAlaGlnArgAspHisLeuAsnValSerLeuValGlyAsnThrValAspL 964
3104 CCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTG 3153
964 euGlyAlaTrpLysTyrLysLeuArgAsnValAsnGlyArgTyrAspLeu 980
3154 CATAATCCG..... 3162
981 TyrAsnProGluValGluLysArgAsnGlnThrValAspThrThrAsnIl 997
3162 3162
997 eThrThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnA 1014
3162 3162
1014 snGluGluIleAlaArgValAspGluAlaProValProProProAlaPro 1030
3162 3162
1031 AlaThrProSerGluThrThrGluThrValAlaGluAsnSerLysGlnGl 1047
3162 3162
1047 uSerLysThrValGluLysAsnGluGlnAspAlaThrGluThrThrAlaG 1064
3163 .....GTCAAAGAACAAGAG 3177
1064 lnAsnArgGluValAlaLysGluAlaLysSerAsnValLysAlaAsnThr 1080
3178 CTTTCCGACAAACTCGGCAAGGCGGGA...GAAACAGAGGCCGCCTTGAC 3224
1081 GlnThrAsnGluValAlaGlnSerGlySerGluThrLysGluThrGlnTh 1097
3225 GGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACG 3274
1097 rThrGluThrLysGluThrAlaThrValGluLysGluGluLysAlaLysV 1114
3275 CGCAAAGC.....CTTGACGCGCTGATTGCG 3300
1114 alGluThrGluLysThrGlnGluValProLysValThrSerGlnValSer 1130
3301 GCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCG 3350
1131 ProLysGlnGluGlnSerGluThrValGlnProGlnAlaGluProAlaAr 1147
3351 GCAGGCAGGCGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAA 3400
1147 g.....GluAsnAspProThrValAsnIleLysGluProGlnS 1160
3401 AACGGGTGCAGGCGGATAAAGACACCGCCTTG...GCGAAACAGCGCGAA 3447
1160 erGlnThrAsnThrThrAlaAspThrGluGlnProAlaLysGluThrSer 1176
3448 GCG..... 3450
1177 SerAsnValGluGlnProValThrGluSerThrThrValAsnThrGlyAs 1193
3451 .....GAAACCCGGCCGGCTACCACCGCCTTCC 3478
1193 nSerValValGluAsnProGluAsnThrThrProAlaThrThrGln...P 1209
3479 CC.....CGCGCCCGCCGC 3492
1209 roThrValAsnSerGluSerSerAsnLysProLysAsnArgHisArgArg 1225
3493 GCCCGCCGGGATTTGCCG...CAACCGCAGCCCCAACCGCAACCCCAACC 3539
1226 SerValArgSerValProHisAsnValGluProAlaThrThrSerSerAs 1242
3540 GCAGCGC.....GACCTGATCAGCCGTTATGCCAATA 3571
1242 nAspArgSerThrValAlaLeuCysAspLeuThrSerThrAsnThrAsnA 1259
3572 GCGGTTTGAGTGAATTTTCCGCC.....ACGCTCAACAGC 3606
1259 laValLeuSerAspAlaArgAlaLysAlaGlnPheValAlaLeuAsnVal 1275
3607 GTTTTCGCCGTACAGGACGAATTGGACCGCGTG...TTTGCCGAAGACCG 3653
1276 GlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluGl 1292
3654 CCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTT 3703
1292 yGlnTyrAsnValTrpValSerAsnThrSerMetAsnLysAsnTyrSerS 1309
3704 CGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGT 3753
1309 erSerGlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGly 1325
3754 ATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTTCGCA 3800
1326 TrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTy 1342
3801 CAACCGGACCGGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGC 3850
1342 rValArgAsnSerAsnAsnPheAspLysAlaThrSerLysAsnThrLeuA 1359
3851 TTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATC 3900
1359 laGlnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTrpTyrLeu 1375
3901 GGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCAT 3950
1376 GlyIleAspLeuGlyTyrGlyLysPheGlnSerLysLeuGlnThrAsnHi 1392
3951 CAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGAT 4000
1392 sAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThrAlaGlyL 1409
4001 ACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACG 4050
1409 ysAlaPheAsnLeuGlyAsnPheGlyIleThrProIleValGlyValArg 1425
4051 CGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGC 4100
1426 TyrSerTyrLeuSerAsnAlaAspPheAlaLeuAspGlnAlaArgIleLy 1442
4101 CACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATT 4150
1442 sValAsnProIleSerValLysThrAlaPheAlaGlnValAspLeuSerT 1459
4151 ATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTG 4200
1459 yrThrTyrHisLeuGlyGlu...PheSerValThrProIleLeuSerAla 1474
4201 TCCTATACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGC 4250
1475 ArgTyr...AspAlaAsnGlnGlySerGlyLysIleAsnValAsnGlyTy 1490
4251 CGTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAA 4300
1490 rAspPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyLeuL 1507
4301 ACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAG 4350
1507 ysLeuLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThrLys 1523

```
    4351 GGGCCGCAATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAGGCTAC 4398
         :::   |||   |||   |||   :::|||   :::||||||::::::
    1524 AlaLysGlnAlaGluLysGlnLysThrAlaGluLeuLysLeuSerPhe 1539

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75564

seq_documentation_block:
ID   AAY75564 standard; Protein; 1431 AA.
XX
AC   AAY75564;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 759 protein sequence SEQ ID NO:2602.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ54326.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 1234; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   1431 AA;


alignment_scores:
           Quality: 1585.50              Length:   1668
             Ratio:  1.829                Gaps:     50
 Percent Similarity:  51.978   Percent Identity: 28.118


alignment_block:
 US-09-303-518D-653 x AAY75564   ..

 Align seg 1/1  to: AAY75564  from: 1  to: 1431

      55 GGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGG 104
         |||   :::     ||||||||||||   :::|||      :::
      17 GlyLeuPheAlaValSerProAlaTyrSerSerIleValArgAsn..... 31

     105 CATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACC 154
                                                  ::::::||||
      32 .....................................AspValAspTyrG 36

     155 AATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCG 204
         |||||:::|||||||||||||||||||||||||   |||:::|||||||||
      36 lnTyrPheArgAspPheAlaGluAsnLysGlyAlaPheThrValGlyAla 52

     205 AAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGAT 254
         ::::::|||::::::   :::|||:::|||::::::::::|||:::   ::
      53 SerAsnIleSerIleGlnAspLysGlnGlyLysIleLeuGlyArgValLe 69

     255 GACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGT...AACGGCG 301
         ::::      ||||||   ||||||   |||   :::|||   :::::::
      69 uAsnGlyIleProMetProAspPheArgValSerAsnArgGlnThrAlaI 86

     302 TGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAACGGC 351
         ::|||:::|||:::      ||||||:::   ||||||   ||||||
      86 leAlaThrLeuValHisProGlnTyrValAsnSerValLysHisAsnVal 102

     352 GGCTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCGATCAGCA 401
         ||||||:::::::::::::||||||   :::   :::||||||::::::::
     103 GlyTyrGlySerIleGlnPheGlyAsnAspThrGlnAsnProGluGluGl 119

     402 CCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTA 451
         :   ::::::|||::::::|||:::|||                    |
     119 nAlaTyrThrTyrArgLeuValSerArg.....................A 129

     452 ACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTT 501
         ||   ||||||         ||||||||||:::||||||||||:::|||:::
     129 snProHisPro...AspTyrAspTyrHisLeuProArgLeuAsnLysLeu 144

     502 GTCACAGATGCAGAACCTGTTGAGATGACCAGT..................... 534
         |||||||:::   :::|||:::   ::::::|||
     145 ValThrGluIleSerProThrAlaLeuSerSerValProLeuLeuGlyAs 161

     535 ....TATATGGATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATC 580
                 :::      |||   |||   ::::::::::|||
     161 nGlyGlnProLysAlaAsnAlaTyrLeuAspThrAspArgPheProTyrP 178

     581 GTGTTCGAATCGGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAA 630
           ||||||:::|||:::|||   |||      |||:::   :::
     178 heValArgLeuGlySerGlyThrGlnGlnValArgLysAlaAspGly... 193

     631 CCCAATAACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGGCT 680
                       :::      :::   |||   ||||||::::::||
     194 ..............ThrArgThrArgThrAlaProAlaTyrGlnTyrLe 205

     681 CGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCA 730
         |:::|||||||:::   :::         |||      :::      :::
     205 uThrGlyGlyThrProLeuLysValLeuGlyPheGlnAsnHisGlyLeuL 222

     731 ACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 780
           :::|||:::   ::::::      :::|||          |||   |||
     222 euValGlyGly...SerLeuThrAspGlnPro.........LeuAsnThr 234

     781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCA 830
           :::      ||||||||||||||||||:::|||   :::|||   ::
     235 TyrAlaIleAlaGlyAspSerGlySerProLeuPheAlaPheAspLysHi 251

     831 AAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATA 880
         ::::::::::|||::::::   ||||||||||:::|||          |||
     251 sGluAsnArgTrpValLeuAlaGlyValLeuSerThr.........TyrA 265
```

```
 881 TAGGAAAAAGCAATGGCTTCCAG......CTAGTTCGTAAAGATTGGTTC 924
 265 laGlyPheAspAsnPhePheAsnLysTyrIleValThrGlnProGluPhe 281

 925 TATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACA 974
 282 IleArgSerThrIleArgGln.................TyrGluThrAr 292

 975 T...............CAAAATGGGAAATACTTTTTTAACGACAATAATA 1009
 292 gLeuAspValGlyLeuThrThrAsnGluLeuIleTrpArgAspAsnGlyA 309

1010 ATGGCGCAGGAAAAATCGATGCCAAACATAAACACTATTCTCTACCTTAT 1059
 309 snGlyAsnSerThrLeuGlnGlyLeuAsnGluArgIleThrLeuPro... 324

1060 AGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGAC 1109
 325 .......................IleAlaAsnProSerLeuAla..... 331

1110 AGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGAC 1159
 332 ..................................................P 332

1160 CCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACAAAGGAAAA... 1206
 332 roGln...AsnAspSerArgHisMetProSerGluAspAlaGlyLysThr 347

1207 ..........................GGTGAATTGATACTTACCAGCAA 1229
 348 LeuIleLeuSerSerArgPheAspAsnLysThrLeuMetLeuAlaAspAs 364

1230 CATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCT 1279
 364 nIleAsnGlnGlyAlaGlyAlaLeuGlnPheAspSerAsnPheThrValV 381

1280 CGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGAT 1329
 381 alGlyLysAsnHis...ThrTrpGlnGlyAlaGlyValIleValAlaAsp 396

1330 GGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTC 1379
 397 GlyLysArgValPheTrpGlnValSerAsnProLysGlyAspArgLeuSe 413

1380 CAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAG 1429
 413 rLysLeuGlyAlaGlyThrLeuIleAlaAsnGlyGlnGlyIleAsnGlnG 430

1430 GCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGGGGAC 1479
 430 lyAspIleSerIleGlyGluGlyThrValValLeuAlaGlnLysAlaAla 446

1480 GATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAG 1529
 447 SerAspGlySerLysGlnAlaPheAsnGlnValGlyIleThrSerGlyAr 463

1530 GGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCT 1579
 463 gGlyThrAlaValLeuAlaAspSerGlnGlnIleLysProGluAsnLeuT 480

1580 ATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCG 1629
 480 yrPheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnAsnLeuAla 496

1630 TTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAA 1679
 497 PheThrHisIleArgHisAlaAspGlyGlyAlaGlnIleValAsnHisAs 513

1680 TCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACTA 1729
 513 nProAspGlnAlaAlaThrLeuThrLeuThrGlyAsnProValLeuSerP 530
```

```
1730 CAACC..................................................... 1734
 530 roGluHisValGluTrpValGlnTrpGlyAsnArgProGlnGlyAsnAla 546

1734 ..................................................... 1734
 547 AlaValTyrGluTyrIleAsnProHisArgAsnArgArgThrAspTyrPh 563

1735 ................GGCAAT.......................... 1740
 563 eIleLeuLysProGlyGlyAsnProArgGluPhePheProLeuAsnMetL 580

1741 ............................AACAACAACTTGGATAGCAAA 1761
 580 ysAsnSerThrSerTrpGlnPheIleGlyAsnAsnArgGlnGlnAlaAla 596

1762 AAAGAAATTGCC..............................TACAA 1778
 597 GluGlnValAlaGlnAlaGluAsnAlaArgProAspLeuIleThrPheGl 613

1779 CGGTTGGTTTGGCGAGAAAGATGCAACC..................... 1806
 613 yGlyTyrLeuGlyGluAsnAlaGlnThrGlyLysAlaAlaProSerTyrS 630

1806 ..................................................... 1806
 630 erLysThrAsnGluAlaAlaIleGluLysThrArgHisIleAlaAsnAla 646

1807 .......................AAAACGAACGGGCGGCTCAATCTGAA 1832
 647 AlaValTyrGlyArgProGluTyrArgTyrAsnGlyAlaLeuAsnLeuHi 663

1833 TTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAA 1882
 663 sTyrArgProLysArgThrAspSerThrLeuLeuLeuAsnGlyGlyMetA 680

1883 ATTTAAACGGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGC 1932
 680 snLeuAsnGlyGluValLeuIleGluGlyGlyAsnMetIleValSerGly 696

1933 AGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAAT 1982
 697 ArgProValProHisAlaTyrAspHisGlnAla..... 707

1983 GGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACC 2032
 708 .............LysArgGluProValLeuGluAsnGluTrpThrAspG 720

2033 GCACATTTAAAGCGGAAAACTTCCATATTCAGGGC...GGACAAGCGGTG 2079
 720 lySerPheLysAlaAlaArgPheThrLeuArgAsnHisAlaArgLeuThr 736

2080 GTTTCCCGCAATGTTGCCAAAGTGGAAGGCGAT.........TGGCATTT 2120
 737 AlaGlyArgAsnThrAlaHisLeuAspGlyAspIleThrAlaTyrAspLe 753

2121 AAGCAAT...CACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCC 2167
 753 uSerGlyIleAspLeuGlyPheThrGlnGlyLysThrProGlu..... 767

2168 ACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAA 2217
 768 ........CysTyrArgSerTyrHisSerGlySerThrHisCysThrPro 781

2218 AAAACCATTACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACAT 2267
 782 AsnAlaValLeuLysAlaGluAsnTyrArgAlaLeuProAlaThrGlnVa 798

2268 CAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGAC 2317
 798 lArgGlyAspIleThrLeuAsnAspArgSerGluLeuArgLeu..... 812

2318 TTGCCACACTCAACGGCAATCTTAGTGCAGGCGGAGACACGCACTATACG 2367
```


813 GlyLysAlaHis 816
2368 GTTACGCGCAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAATGC 2417
816 816
2418 CCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGG 2467
816 816
2468 ACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCTG 2517
816 816
2518 ACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGG 2567
816 816
2568 CAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCT 2617
817 L 817
2618 TTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAA... 2664
817 euTyrGlySerIleArgAlaGlyLysAspThrAlaValArgMetGluAla 833
2665 GACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAACCT 2714
834 AspSerAsnTrpThrLeuSerGlnSerSerHisThrGlyAlaLeuThrLe 850
2715 TGACAACGCCACCATTACACTCAATTCCGCCTATCGACACGATGCGGCAG 2764
850 uAspGlyAlaGlnIleThrLeuAsnProAspPheAlaAsnAsn..... 864
2765 GCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGT 2814
864 864
2815 TCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACAC 2864
865 ThrHisAsnAsnArgPheAsnTh 872
2865 GCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGT 2914
872 rLeuThrValAsnGlyThrLeuAspGlyPheGlyThrPheArgPheLeuT 889
2915 CGGAACTCTTCGGCTACCGCAGCGGC...AAATTGAAGCTGGCGGAAAGT 2961
889 hrGlyIleValArgLysGlnAsnAlaProProLeuLysLeuGluGlyAsp 905
2962 TCCGAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAACCCGT 3011
906 SerArgGlyAlaPheGlnIleHisValLysAsnThrGlyGlnGluProGl 922
3012 AAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGT 3061
922 nThrThrGluSerLeuAlaLeuVal.....SerLeuAsnProLysHisS 937
3062 CCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCA 3111
937 erHisGlnAlaArgPheThrLeuGlnAsnGlyTyrAlaAspLeuGlyAla 953
3112 TGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCC 3161
954 TyrArgTyrIleLeuArgLysAsnAsnAsnGlyTyrSerLeuTyrAsnPr 970
3162 GGTCAAAGAACAAGAGCTT..... 3180
970 oLeuLysGluAlaGluLeuGlnIleGluAlaThrArgAlaGluHisGluA 987
3180 3180


987 rgAsnGlnGlnAlaTyrAsnGlnLeuGlnAlaThrAspIleSerArgGln 1003
3181 .....TCCGACAAACTCGGCAAGGCG..... 3201
1004 ValGlnHisAspSerAspAlaThrArgGlnAlaLeuGlnAlaTrpGlnAs 1020
3202 .GGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCCA 3250
1020 nSerGlnThrGluLeuAlaArgIleAspSerGlnValGlnTyrLeuSerA 1037
3251 AACAACAGGCGGAAAAAGAC.....AACGCGCAA 3279
1037 laGlnLeuLysGlnThrAspProLeuThrGlyIleLeuThrArgAlaGln 1053
3280 AGCCTTGACGCGCTGATTGCGGCCGGGCGCAATGCC..... 3315
1054 AsnLeuCysAla.....AlaGlnGlyTyrSerAlaAspIleCysArgGl 1068
3315 3315
1068 nValAlaLysAlaAlaAspThrAsnAspLeuThrLeuPheGluThrGluL 1085
3316 .....ACCGAAAAGGCAGAAAGTGTTGCCGAACCG 3345
1085 euAspThrTyrIleGluArgValGluMetAlaGluSerGluLeuAspLys 1101
3346 GCCCGGCAGGCAGGCGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGA 3395
1102 AlaArgGlnGlyGlyAsp.....AlaGlnAlaValGluTh 1113
3396 GAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAG...C 3442
1113 rAlaArgHisAlaTyrLeuAsnAlaLeuAsnArgLeuSerArgGlnIleH 1130
3443 GCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGC 3492
1130 isSerLeuLysThrGlyValAlaGlyIleArgMetProAsnLeu..... 1144
3493 GCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCA 3542
1144 1144
3543 GCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCG 3592
1145 .AlaGluLeuIleSerArgSerAlaAsnThrAlaValSerGluGlnAlaA 1161
3593 CCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTT 3642
1161 laTyrAsnThrGlyArgGlnGlnAlaGlyArgArgIleAspArgHisLeu 1177
3643 GCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAA 3692
1178 ThrAspProGlnGlnGlnAsnIleTrpLeuGluThrGlyThrGlnGlnTh 1194
3693 ACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGC 3742
1194 rAspTyrHisSerGlyThrHisArgProTyrGlnGlnThrThrAsnTyrA 1211
3743 GCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATC 3789
1211 laHisIleGlyIleGlnThrGlyIleThrAspArgLeuSerValGlyThr 1227
3790 CTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACGGCATCGGC.. 3837
1228 IleLeuThrAspGluArgThrAsnAsnArgPheAspGluGlyValSerAl 1244
3838 ....AACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCA 3883
1244 aArgAsnArgSerAsnGlyAlaHisLeuPheValLysGlyGluAsnGlyA 1261
3884 TCGGCAGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGC 3933
1261 la.....LeuPheAlaAlaAlaAspLeuGlyTyrSerAsnSer 1273

```
3934 ......AGCCTTTCAGACGGCATCAGAGGCAAAATCCGCCGCCGCCGCGTGCT 3977
1274 ArgThrArgPheThrAspTyrAspGlyAlaAlaValArgArgHisAlaTr 1290

3978 GCATTACGGCATTCAGGCAAGATACCGCGCAGGTTTCGGCGGATTCGGCA 4027
1290 pAspAlaGlyIleAsnThrGlyIleLysIleAsp...ThrGlyIleAsnL 1306

4028 TCGAACCGCACATCGGC.....................GCAACGCGCTAT 4056
1306 euArgProTyrAlaGlyIleArgIleAsnArgSerAsnGlyAsnArgTyr 1322

4057 TTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGCCACCCC 4106
1323 ValLeuAspGlyAlaGlu.....................IleAsnSerPr 1332

4107 GGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCAT 4156
1332 oAlaGlnIleGlnThrThrTrpHisAlaGlyIleArgLeuAspLysThrV 1349

4157 TCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTAT 4206
1349 alGluLeuGlyGln...AlaLysLeuThrProAlaPheSerSerAspTyr 1364

4207 ......ACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGC 4250
1365 TyrHisThrArgGlnAsnSerGlySerAlaLeuSer...ValAsnAspAr 1380

4251 CGTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAA 4300
1380 gThrLeuLeuGlnGlnAlaAlaHisGlyThrLeuHisThrLeuGlnIleA 1397

4301 ACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAG 4350
1397 spAlaGlyTyrLysGlyTrpAsnAlaLysLeuHisAlaAlaTyrGlyLys 1413

4351 GGGCCGCAATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAGGCTACCG 4400
1414 AspSerAsnThrAlaArgHisLysGlnAlaGlyIleLysIleGlyTyrAs 1430

4401 CTGG 4404
1430 nTrp 1431
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB52592

seq_documentation_block:

```
ID   ABB52592 standard; Protein; 1376 AA.
XX
AC   ABB52592;
XX
DT   11-FEB-2002  (first entry)
XX
DE   Escherichia coli polypeptide SEQ ID NO 560.
XX
KW   Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW   immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW   systemic infection; non-diarrhoeal infection; septicaemia;
KW   pyelonephritis; antibiotic resistance.
XX
OS   Escherichia coli.
XX
PN   WO200166572-A2.
XX
PD   13-SEP-2001.
XX
PF   12-MAR-2001; 2001WO-EP03445.
XX
PR   10-MAR-2000; 2000FR-0003145.
PR   02-FEB-2001; 2001FR-0001449.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Bingen E,  Bonacorsi S,  Clermont O,  Nassif X,  Tinsley C;
XX
DR   WPI; 2001-550253/61.
XX
PT   A library of DNA fragments of Escherichia coli strains for the
PT   phylogenic determination of a given strain comprises polynucleotides of
PT   nature B2/D+ A- -
XX
PS   Example 6; Fig 6; 646pp; English.
XX
CC   The invention relates to a library of DNA fragments of Escherichia coli
CC   strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
CC   and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
CC   B2/D+A-. The polynucleotides have potential antiinflammatory,
CC   antibacterial and immunosuppressive activity as part of pharmaceutical
CC   compositions used to treat, palliate or prevent extra-intestinal E. coli
CC   infections. The polypeptides are useful for determining the phylogenic
CC   group of a given E. coli strain. These polypeptides can detect and treat
CC   an undesired development of E. coli, particularly an extra-intestinal
CC   infection that include systemic and non-diarrhoeal infections such as
CC   septicaemia, pyelonephritis and meningitis this is particularly
CC   advantageous as bacterial resistance is increasing with the more
CC   frequent use of broad spectrum antibiotics.
XX
SQ   Sequence   1376 AA;
```

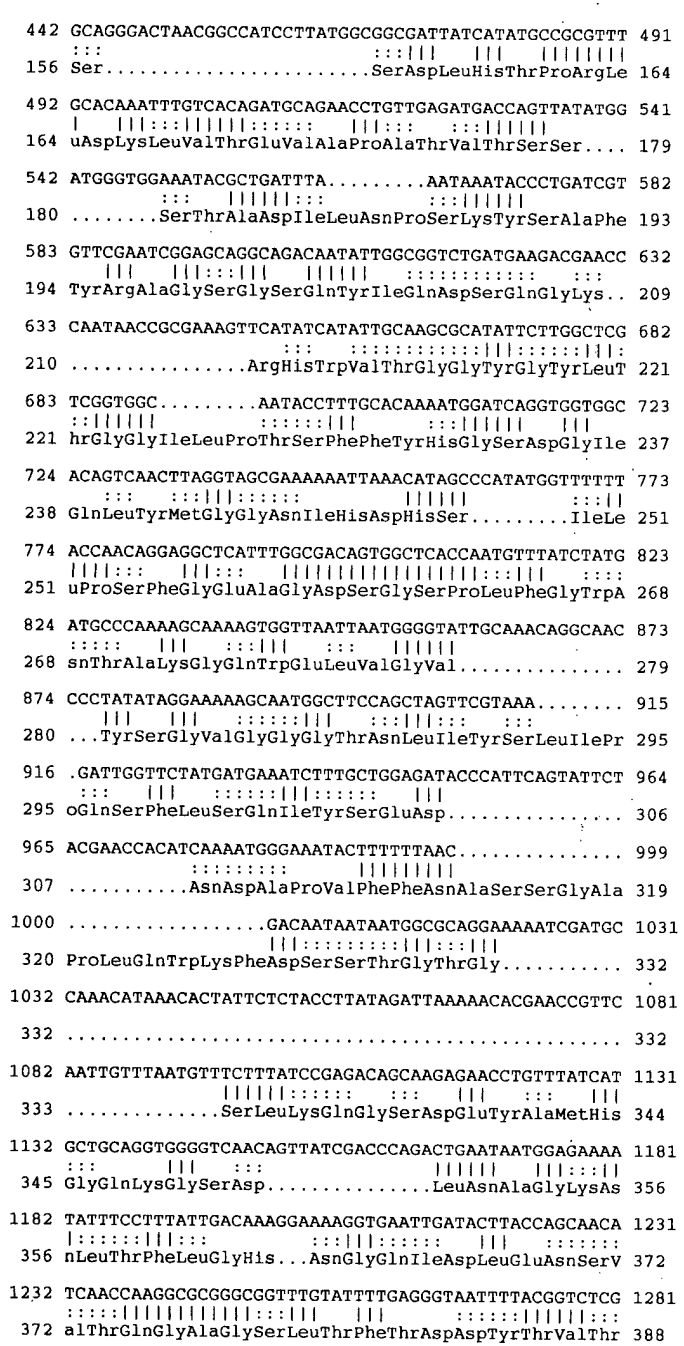
alignment_scores:

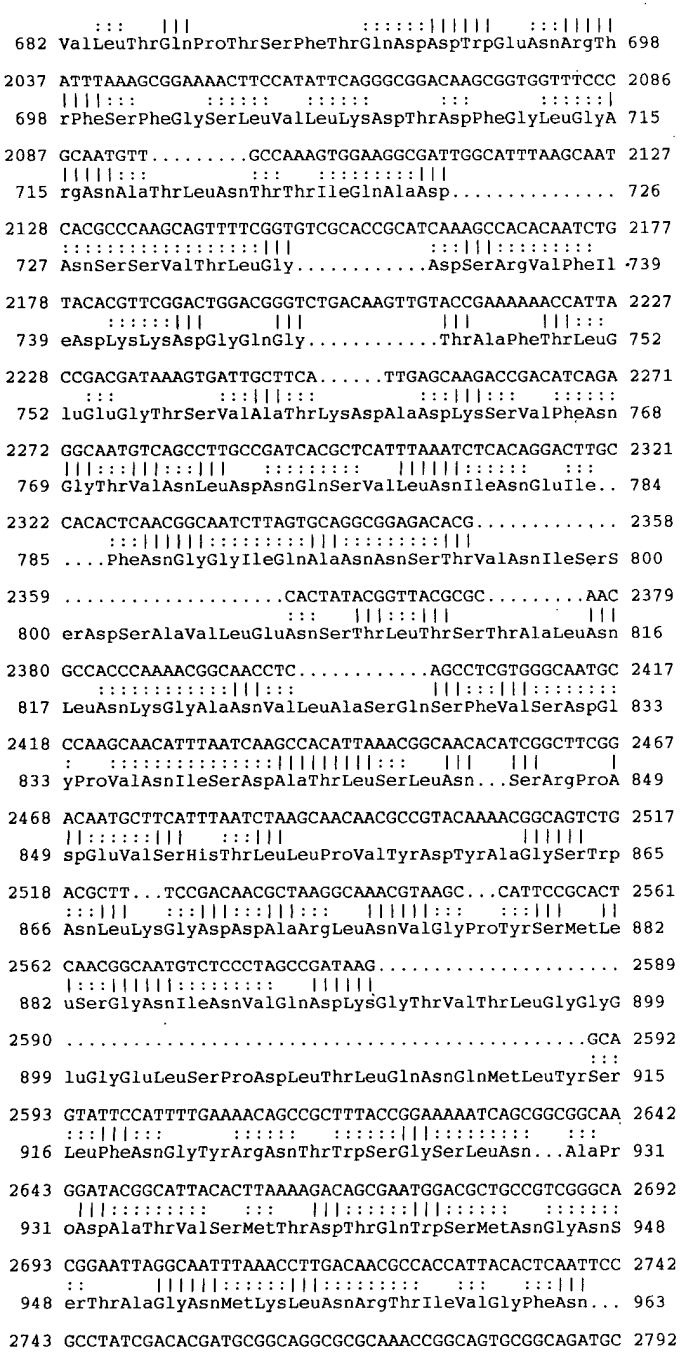
Quality: 1052.50 Length: 1562
Ratio: 1.260 Gaps: 71
Percent Similarity: 53.457 Percent Identity: 25.928

alignment_block:
US-09-303-518D-653 x ABB52592 ..

Align seg 1/1 to: ABB52592 from: 1 to: 1376

```
  10 ACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCTAAA......AC 53
  15 ThrValLysValValSerGluLeuAlaArgArgValCysLysGlySerTh 31

  54 CGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCG 103
  31 rArgArgGlyLysArgLeuSerValLeuThrSerLeuAlaLeuSer...A 47

 104 GCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATCAAC... 150
  47 laLeuLeuProThrValAlaGlyAlaSerThrValGlyGlyAsnAsnPro 63

 151 TACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGG 200
  64 TyrGlnThrTyrArgAspPheAlaGluAsnLysGlyGlnPheGlnAlaGl 80

 201 GGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAAT 250
  80 yAlaThrAsnIleProIlePheAsnAsnLysGlyGluLeuValGly...H 96

 251 CGATGACGAAAGCCCCGATGATTGATTTTTCT.........GTGGTATCG 291
  96 isLeuAspLysAlaProMetValAspPheSerSerValAsnValSerSer 112

 292 CGTAACGGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGC 341
 113 AsnProGlyValAlaThrLeuIleAsnProGlnTyrIleAlaSerValLy 129

 342 ACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATC 391
 129 sHisAsnLysGlyTyrGlnSerValSerPheGly...AspGlyGlnAsn. 144

 392 CCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAA 441
 145 .................SerTyrHisIleValAspArgAsnGluHisSer 155
```




682 ValLeuThrGlnProThrSerPheThrGlnAspAspTrpGluAsnArgTh 698
2037 ATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTTCCC 2086
698 rPheSerPheGlySerLeuValLeuLysAspThrAspPheGlyLeuGlyA 715
2087 GCAATGTT.....GCCAAAGTGGAAGGCGATTGGCATTTAAGCAAT 2127
715 rgAsnAlaThrLeuAsnThrThrIleGlnAlaAsp..... 726
2128 CACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTG 2177
727 AsnSerSerValThrLeuGly.....AspSerArgValPheIl 739
2178 TACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTA 2227
739 eAspLysLysAspGlyGlnGly.....ThrAlaPheThrLeuG 752
2228 CCGACGATAAAGTGATTGCTTCA.....TTGAGCAAGACCGACATCAGA 2271
752 luGluGlyThrSerValAlaThrLysAspAlaAspLysSerValPheAsn 768
2272 GGCAATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGC 2321
769 GlyThrValAsnLeuAspAsnGlnSerValLeuAsnIleAsnGluIle.. 784
2322 CACACTCAACGGCAATCTTAGTGCAGGCGGAGACACG..... 2358
785 ....PheAsnGlyGlyIleGlnAlaAsnAsnSerThrValAsnIleSerS 800
2359 .....CACTATACGGTTACGCGC.....AAC 2379
800 erAspSerAlaValLeuGluAsnSerThrLeuThrSerThrAlaLeuAsn 816
2380 GCCACCCAAAACGGCAACCTC.....AGCCTCGTGGGCAATGC 2417
817 LeuAsnLysGlyAlaAsnValLeuAlaSerGlnSerPheValSerAspGl 833
2418 CCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGG 2467
833 yProValAsnIleSerAspAlaThrLeuSerLeuAsn...SerArgProA 849
2468 ACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCTG 2517
849 spGluValSerHisThrLeuLeuProValTyrAspTyrAlaGlySerTrp 865
2518 ACGCTT...TCCGACAACGCTAAGGCAAACGTAAGC...CATTCCGCACT 2561
866 AsnLeuLysGlyAspAspAlaArgLeuAsnValGlyProTyrSerMetLe 882
2562 CAACGGCAATGTCTCCCTAGCCGATAAG..... 2589
882 uSerGlyAsnIleAsnValGlnAspLysGlyThrValThrLeuGlyGlyG 899
2590 .....GCA 2592
899 luGlyGluLeuSerProAspLeuThrLeuGlnAsnGlnMetLeuTyrSer 915
2593 GTATTCCATTTTGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAA 2642
916 LeuPheAsnGlyTyrArgAsnThrTrpSerGlySerLeuAsn...AlaPr 931
2643 GGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCA 2692
931 oAspAlaThrValSerMetThrAspThrGlnTrpSerMetAsnGlyAsnS 948
2693 CGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCC 2742
948 erThrAlaGlyAsnMetLysLeuAsnArgThrIleValGlyPheAsn... 963
2743 GCCTATCGACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGC 2792
963 963
2793 GCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTT 2842
964 .....G 964
2843 CGGCAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGT 2892
964 lyGlyThrSerSerPheThrThrLeuThrThrAspAsnLeuAspAlaVal 980
2893 CAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAA 2942
981 GlnSerAlaPheValMetArgThrAspLeu.....AsnLysAlaAspLy 995
2943 ATTGAAGCTGGCGGAAAGTTCCGAAGGC...ACTTACACCTTGGCTGTCA 2989
995 sLeuValIleAsnLysSerAlaThrGlyHisAspAsnSerIleTrpValA 1012
2990 AC.....AATACCGGCAACGAACCC...GTA 3012
1012 snPheLeuLysLysProSerAspLysAspThrLeuAspIleProLeuVal 1028
3013 AGTCTCGAGCAATTGACG.....GT 3032
1029 SerAlaProGluAlaThrAlaAspAsnLeuPheArgAlaSerThrArgVa 1045
3033 AGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCACCC 3082
1045 lValGlyPheSerAspValThrProThrLeuSerValArgLysGluAspG 1062
3083 TGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGC 3132
1062 lyLysLysGluTrpValLeuAspGly.....TyrGlnValAlaArg 1075
3133 AAAGACGGCGAG.....TTCCGCCTGCA 3155
1076 AsnAspGlyGlnGlyLysAlaAlaAlaThrPheMetHisIleSerTyrAs 1092
3156 TAATCCGGTCAAAGAA.....CAAGAGCTTT 3181
1092 nAsnPheIleThrGluValAsnAsnLeuAsnLysArgMetGlyAspLeuA 1109
3182 CCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAA 3231
1109 rgAspIleAsnGlyGluAlaGlyThrTrpValArgLeuLeuAsnGlySer 1125
3232 CAGGCA.....CAACTTGC 3245
1126 GlySerAlaAspGlyGlyPheThrAspHisTyrThrLeuLeuGlnMetGl 1142
3246 CGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGA 3295
1142 yAlaAspArgLysHisGluLeuGlySerMetAspLeuPheThrGlyValM 1159
3296 TTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCG 3345
1159 etAlaThrTyrThrAspThrAspAlaSerAlaGlyLeuTyrSerGlyLys 1175
3346 GCCCGGCAGGCAGGCGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGA 3395
1176 ThrLysSerTrpGlyGly.....GlyPheTyrAlaSerGlyLeuPh 1189
3396 GAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCG 3445
1189 eArgSerGlyAlaTyrPheAsp.....LeuIleAlaLysTyrIleH 1203
3446 AAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCC 3495
1203 isAsnGluAsnLys..... 1207
3496 CGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCG 3545
1208 .....Ty 1208

```
3546 CGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCA 3595
1208 rAspLeu......AsnPheAlaGlyAlaGlyLysGlnAsnPheArgSer. 1222

3596 CGCTCAACAGCGTTTTCGCCGTACAGGACGAA..................... 3627
1223 .....HisSerLeuTyrAlaGlyAlaGluValGlyTyrArgTyrHisLeu 1237

3628 TTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGG 3677
1238 ThrAspThrThrPheValGluProGlnAlaGluLeuValTrp........ 1251

3678 CATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCC 3727
1252 .............GlyArgLeuGlnGlyGlnThrPheAsnTrpAsnAspS 1264

3728 AACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGC... 3774
1264 erGlyMetAsp.........ValSerMetArgArgAsnSerValAsnPro 1277

3775 ......GGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACAC 3818
1278 LeuValGlyArgThrGlyValVal..............SerGlyLysTh 1289

3819 CTTC.........GACGACGGCATCGGCAACTCGGCACGGCTTGCCCACG 3859
1289 rPheSerGlyLysAspTrpSerLeuThrAlaArgAlaGlyLeuHisTyr. 1305

3860 GTGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATCGGCATCAGC 3909
1306 ............................GluPheAspLeuThrAspSer 1312

3910 GCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGAGGCAA 3959
1313 AlaAspValHisLeuLysAspAlaAlaGlyGluHisGlnIleAsnGlyAr 1329

3960 AATCCGCCGCCGCGTGCTGCAT...TACGGCATTCAGGCAAGATAC.... 4002
1329 gLysAspGlyArgMetLeuTyrGlyValGlyLeuAsnAlaArgPheGlyA 1346

4003 ............................CGCGCAGGTTTCGGCGGATTC 4023
1346 spAsnThrArgLeuGlyLeuGluValGluArgSerAlaPheGlyLysTyr 1362

4024 GGCATCGAACCGCACATCGGCGCAACG...CGCTAT 4056
1363 AsnThrAspAspAlaIleAsnAlaAsnIleArgTyr 1374
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB52745

seq_documentation_block:

```
ID   ABB52745 standard; Protein; 1371 AA.
XX
AC   ABB52745;
XX
DT   11-FEB-2002  (first entry)
XX
DE   Escherichia coli polypeptide SEQ ID NO 891.
XX
KW   Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW   immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW   systemic infection; non-diarrhoeal infection; septicaemia;
KW   pyelonephritis; antibiotic resistance.
XX
OS   Escherichia coli.
XX
PN   WO200166572-A2.
XX
PD   13-SEP-2001.
XX
PF   12-MAR-2001; 2001WO-EP03445.
XX
PR   10-MAR-2000; 2000FR-0003145.
PR   02-FEB-2001; 2001FR-0001449.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Bingen E,  Bonacorsi S,  Clermont O,  Nassif X,  Tinsley C;
XX
DR   WPI; 2001-550253/61.
XX
PT   A library of DNA fragments of Escherichia coli strains for the
PT   phylogenic determination of a given strain comprises polynucleotides of
PT   nature B2/D+ A- -
XX
PS   Example 6; Fig 6; 646pp; English.
XX
CC   The invention relates to a library of DNA fragments of Escherichia coli
CC   strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
CC   and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
CC   B2/D+A-. The polynucleotides have potential antiinflammatory,
CC   antibacterial and immunosuppressive activity as part of pharmaceutical
CC   compositions used to treat, palliate or prevent extra-intestinal E. coli
CC   infections. The polypeptides are useful for determining the phylogenic
CC   group of a given E. coli strain. These polypeptides can detect and treat
CC   an undesired development of E. coli, particularly an extra-intestinal
CC   infection that include systemic and non-diarrhoeal infections such as
CC   septicaemia, pyelonephritis and meningitis this is particularly
CC   advantageous as bacterial resistance is increasing with the more
CC   frequent use of broad spectrum antibiotics.
XX
SQ   Sequence   1371 AA;
```

alignment_scores:

Quality: 1048.00 Length: 1570
Ratio: 1.340 Gaps: 69
Percent Similarity: 49.809 Percent Identity: 25.796

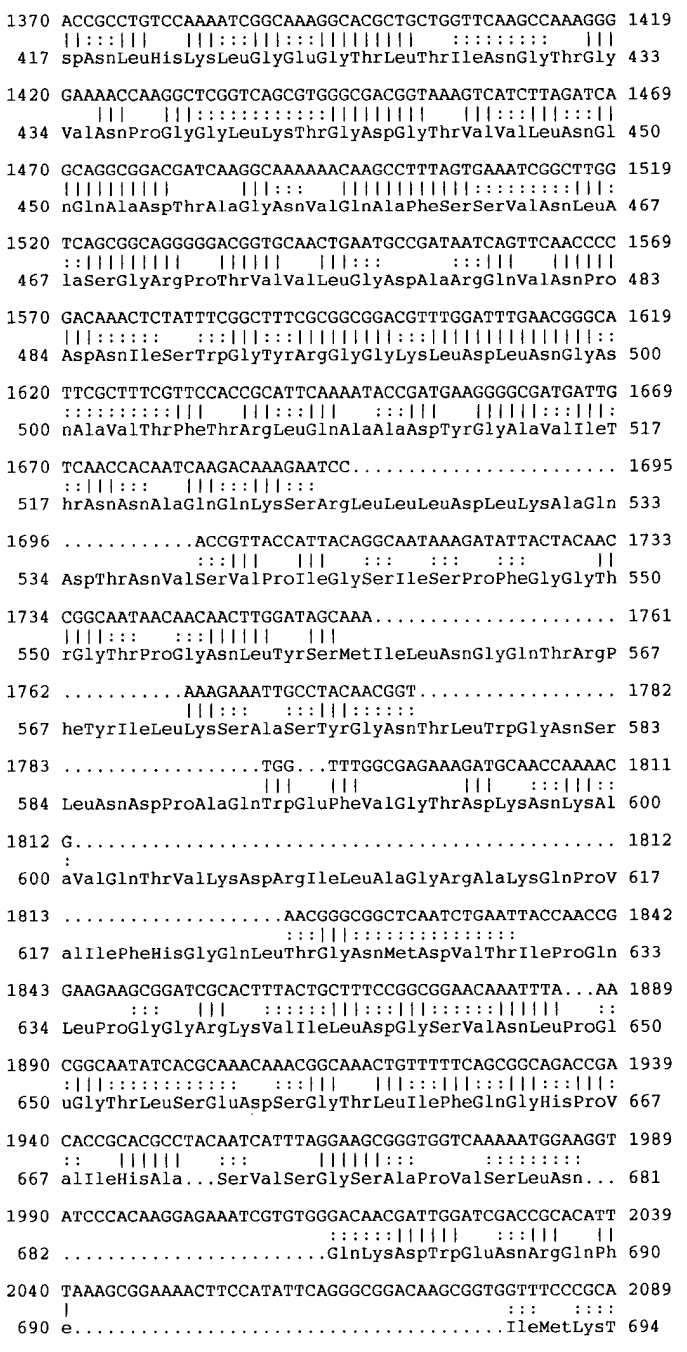
alignment_block:

US-09-303-518D-653 x ABB52745 ..

Align seg 1/1 to: ABB52745 from: 1 to: 1371

```
112 CCCCAA...GCCCGGGCGGGACACACTTATTTCGGCATCAACTACCAATA 158
 50 ProGlnIleSerGlnAlaGlyIleValArgSerAspIleAlaTyrGlnIl 66

159 CTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAG 208
 66 eTyrArgAspPheAlaGluAsnLysGlyLeuPheValProGlyAlaThrA 83

209 ATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACG 258
 83 spIleProValTyrAspLysAspGlyLysLeuValGlyArg...LeuAsp 98

259 AAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAACGGCGTGGCGGC 308
 99 LysAlaProMetAlaAspPheSerSerValSerSerAsnGlyValAlaTh 115

309 ATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATA 358
115 rLeuValSerProGlnTyrIleValSerValLysHisAsnGlyGlyTyrG 132

359 ACAATGTTGATTTTGGTGCCGGAGGGAAGCAATCCCGATCAGCACCGCTTT 408
132 lnSerValSerPheGly...AsnGlyLysAsn..................... 141

409 TCTTACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCA 458
142 ThrTyrSerLeuValAspArgAsnAsnHisSerSer.............. 153

459 TCCTTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAG 508
154 .........ValAspPheHisAlaProArgLeuAsnLysLeuValThrG 167
```


509 ATGCAGAACCTGTTGAGATGACCAGTTATATGGATGGGTGGAAA..... 552
167 luValIleProSerAlaIleThrSer.....GluGlyThrLysAlaAsn 181
553 ...TACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATCGGAGCAGG 599
182 AlaTyrLysAspThrGluArgTyrThrAlaPheTyrArgValGlySerGl 198
600 CAGACAATATTGGCGGTCTGATGAAGACGAACCCAATAACCGCGAAAGTT 649
198 yThrGlnTyrThrLys...AspLysAsp.....GlyAsnL 209
650 CATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAATACCTTT 699
209 euValLysValAlaGlyGlyTyrAlaPheLysThrGlyGlyThrThr... 224
700 GCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAAAAT 749
225 .....GlyValProLeuIleSerAspAlaThrIl 234
750 TAAACATAGCCCA.....TATGGTTTTTTACCAA 778
234 eValSerAsnProGlyGlnThrTyrAsnProValAsnGlyProLeuProA 251
779 CAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCC 828
251 spTyrGlyAlaProGlyAspSerGlySerProLeuPheAlaTyrAspGlu 267
829 CAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTA 878
268 GlnGlnLysLysTrpValIleValAlaValLeuArgAlaTyrAlaGlyIl 284
879 TATAGGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTC.... 924
284 eAsnGlyAlaThrAsnTrpTrpAsnValIleProThrAspTyrLeuAsnG 301
925 .....TATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAA 969
301 lnValMetGlnAspAspPheAspAlaProValAspPheValSerGlyLeu 317
970 CCACATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGG 1019
318 ProProLeuAsnTrpThrTyr.....AspLysThrSerGlyThrGl 331
1020 AAAAATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAA 1069
331 yThrLeuSerGlnGlySerLysAsnTrpThrMet..... 342
1070 CACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAA 1119
342 342
1120 CCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAA 1169
343 .....HisGlyGlnLysAspAsnAsp.....LeuAs 351
1170 TAATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATAC 1219
351 nAlaGlyLysAsnLeuValPheSerGlyGln...AsnGlyAlaIleValL 367
1220 TTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAAT 1269
367 euLysAspSerValThrGlnGlyAlaGlyTyrLeuGluPheLysAspSer 383
1270 TTTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCA 1319
384 TyrThrValSerAlaGluSerGlyLysThrTrpThrGlyAlaGlyIleIl 400
1320 TATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACG 1369
400 eThrAspLysGlyThrAsnValThrTrpLysValAsnGlyValAlaGlyA 417


1370 ACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGG 1419
417 spAsnLeuHisLysLeuGlyGluGlyThrLeuThrIleAsnGlyThrGly 433
1420 GAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCA 1469
434 ValAsnProGlyGlyLeuLysThrGlyAspGlyThrValValLeuAsnGl 450
1470 GCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGG 1519
450 nGlnAlaAspThrAlaGlyAsnValGlnAlaPheSerSerValAsnLeuA 467
1520 TCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCC 1569
467 laSerGlyArgProThrValValLeuGlyAspAlaArgGlnValAsnPro 483
1570 GACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCA 1619
484 AspAsnIleSerTrpGlyTyrArgGlyGlyLysLeuAspLeuAsnGlyAs 500
1620 TTCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTG 1669
500 nAlaValThrPheThrArgLeuGlnAlaAlaAspTyrGlyAlaValIleT 517
1670 TCAACCACAATCAAGACAAAGAATCC..... 1695
517 hrAsnAsnAlaGlnGlnLysSerArgLeuLeuLeuAspLeuLysAlaGln 533
1696 .....ACCGTTACCATTACAGGCAATAAAGATATTACTACAAC 1733
534 AspThrAsnValSerValProIleGlySerIleSerProPheGlyGlyTh 550
1734 CGGCAATAACAACAACTTGGATAGCAAA..... 1761
550 rGlyThrProGlyAsnLeuTyrSerMetIleLeuAsnGlyGlnThrArgP 567
1762 .....AAAGAAATTGCCTACAACGGT..... 1782
567 heTyrIleLeuLysSerAlaSerTyrGlyAsnThrLeuTrpGlyAsnSer 583
1783 .....TGG...TTTGGCGAGAAAGATGCAACCAAAAC 1811
584 LeuAsnAspProAlaGlnTrpGluPheValGlyThrAspLysAsnLysAl 600
1812 G..... 1812
600 aValGlnThrValLysAspArgIleLeuAlaGlyArgAlaLysGlnProV 617
1813 .....AACGGGCGGCTCAATCTGAATTACCAACCG 1842
617 alIlePheHisGlyGlnLeuThrGlyAsnMetAspValThrIleProGln 633
1843 GAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAATTTA...AA 1889
634 LeuProGlyGlyArgLysValIleLeuAspGlySerValAsnLeuProGl 650
1890 CGGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGA 1939
650 uGlyThrLeuSerGluAspSerGlyThrLeuIlePheGlnGlyHisProV 667
1940 CACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGT 1989
667 alIleHisAla...SerValSerGlySerAlaProValSerLeuAsn... 681
1990 ATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACATT 2039
682 .....GlnLysAspTrpGluAsnArgGlnPh 690
2040 TAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTTCCCGCA 2089
690 e.....IleMetLysT 694
2090 ATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCA 2139

```
         ::::::::         ::::::|||:::||||||||||::::::|||:::
 694 hrLeuSerLeuLysAspAlaAspPheHisLeuSerArgAsnAlaSerLeu 710

2140 GTTTTCGGTGTCGCACCGCATCAAAGCCAC.................... 2169
                   :::        :::||||||
 711 AsnSerAspIleLysSerAspAsnSerHisIleThrLeuGlySerAspAr 727

2170 .ACAATCTGTACACGTTCGGACTGGACGGGT...CTGACAAGTTGTACCG 2215
      ::::::      :::::|||  ||||||       :::          |
 727 gValPheValAspLysAsnAspGlyThrGlyAsnTyrValIleLeuGluG 744

2216 AAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGAC 2265
     ||   |||   :::   |||   |||            ::::::::::::
 744 luGlyThrSerValProAspThrVal.........AsnAspArgSerGln 757

2266 ATCAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAATCTC..... 2310
        :::||||||::::::|||        ::::::   |||::::::
 758 TyrGluGlyAsnIleThrLeuAspHisAsnSerThrLeuAspIleGlySe 774

2311 .......ACAGGA..................................... 2316
            ||||||
 774 rArgPheThrGlyGlyIleGluAlaTyrAspSerAlaValSerIleThrS 791

2316 .................................................. 2316

 791 erProAspValLeuLeuThrAlaProGlyAlaPheAlaGlySerSerLeu 807

2317 ....................CTTGCCACACTCAACGGCAATCTTAGTGC 2345
                         |||::::::|||||||||    :::|||
 808 ThrValHisAspGlyGlyHisLeuThrAlaLeuAsnGlyLeuPheSer.. 823

2346 AGGCGGAGACACGCACTATACGGTTACGCGCAACGCCACCCAAAACGGCA 2395
             |||   |||                 :::|||   :::|||:::
 824 .......AspGlyHisIle............GlnAlaGlyLysAsnSerL 834

2396 ACCTCAGCCTCGTGGGCAAT...GCCCAAGCAACATTTAATCAAGCCACA 2442
     ::::::::|||  |||:::   ::::::   |||   ||||||   :::
 834 ysIleThrLeuSerGlyThrProValLysAspThrAlaAsnGlnTyrAla 850

2443 ...............TTAAACGGCAACACATCGGCTTCGGACAATGCTTC 2477
                    :::|||         ::::::||||||||||::
 851 ProAlaValTyrLeuThrAspGlyTyrAspLeuThrGlyAspAsnAlaTh 867

2478 ATTTAATCTAAGCAACAACGCC............................ 2499
     ::::::::::::::::::|||
 867 rLeuGluIleThrArgGlyAlaHisAlaSerGlyAspIleHisAlaSerA 884

2500 ..........GTACAAAACGGCAGTCTGACGCTTTCCGACAACGCTAAG 2538
               |||      ||||||   |||   ::::::   |||:::
 884 laAlaSerThrValThrIleGlySerAspThrProAlaGluLeuAlaSer 900

2539 GCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAA 2588
     |||:::::::::   |||||||:::   |||
 901 AlaGluThrThrAlaSerAlaPheAlaGly.................... 910

2589 GGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCG 2638
      ::::::::::::        |||:::   |||:::|||   |||:::||||
 911 .SerLeuLeuGluGlyTyrAsnAlaAlaPheAsnGlyAlaIleThrGlyG 927

2639 GCAAG...GATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCG 2685
     ||:::   |||:::::::::|||         ::::::   |||||||||
 927 lyArgAlaAspValSerMetHis......AsnAlaLeuTrpThrLeuGly 941

2686 TCGGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACT 2735
     :::   :::   :::   :::|||::::::   |||:::   |||:::
 942 GlyAspSerAlaIleHisThrLeuThrValArgAsnSerArgIleSer.. 957

2736 CAATTCCGCCTATCGACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGG 2785

 957 .................................................. 957

2786 CAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCG 2835

 957 .................................................. 957

2836 CCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATT 2885
        ::::::::::::        |||:::||||||||||||||||   |||||
 958 ...SerGluGlyAspArgThrPheArgThrLeuThrValAsn...LysLe 972

2886 GAACGGTCAGGGAACA...TTCCGCTTTATGTCGGAACTCTTCGGCTACC 2932
     |::::::   |||:::   |||   :::   ::::::|||         :
 972 uAspAlaThrGlySerAspPheValLeuArgThrAspLeu......LysA 987

2933 GCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTG 2982
     :::::   |||::::::::::::|||::::::   |||:::
 987 snAlaAspLysIleAsnValThrGluLysAlaThrGlySer......... 1000

2983 GCTGTCAACAATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGT 3032

1000 .................................................. 1000

3033 AGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCACCC 3082
                  ||||||         |||   |||:::::::|||
1001 .............AspAsn.........SerLeuAsnValSerPhe.... 1008

3083 TGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGC 3132

1008 .................................................. 1008

3133 AAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTC 3182
                       :::   :::|||::::::   |||:::|||::
1009 ..................MetLysAspProAlaGlnGlyGlnSerLeuAs 1019

3183 CGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAAC 3232
       :      |||      |||         |||:::   :::|||::::
1019 nIleProLeuValThrAla.........ProAlaGlyThrSerAlaGluM 1033

3233 AGGCACAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGC 3282
      ::   :::   :::::::::
1033 etPheLysAlaGlyThrArg.............................. 1039

3283 CTTGACGCGCTGATTGCGGCCGGGCGC...AATGCCACCGAAAAGGCAGA 3329
                :::|||:::   :::|||   :::   |||      :::::
1040 .........MetIleGlyPheSerArgValThrProThrLeuHisValAs 1053

3330 AAGTGTTGCCGAACCGGCCCGGCAGGCAGGCGGGAAAATGCCGGCATTA 3379
     ::::                     :::||||||::::::      |||:
1053 pThr.....................SerGlyGlyAsnThrLysTrpIleL 1063

3380 TGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCC 3429
     ::::::::                    :::|||:::   |||   |||
1063 euAspGly.................PheLysAlaGluAlaAspLysAla 1073

3430 TTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCC 3479
        |||   :::   ::::::
1074 AlaAlaAlaLysAlaAspSer............................ 1080

3480 CCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGC 3529

1080 .................................................. 1080

3530 AACCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTG 3579
                                         :::   |||:::|||
1081 ................................PheMetAsnAlaGlyTyr 1086

3580 AGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATT 3629
     :::::::|||   :::   :::|||::::::
1087 LysAsnPheMetThrGluValAsnAsnLeu.................... 1096

3630 GGACCGCGTGTTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCA 3679
                       ::::::|||            :::   :
1097 ..................AsnLysArg...........MetGlyAspL 1103

3680 TCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAA 3729
     ::|||||||||            :::   |||         |||:::   :::
1103 euArgAspThr............AsnGlyAspAlaGlyAlaTrpAlaArg 1115
```

```
3730 .........CAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGG 3770
                :::       :::   |||   :::   |||
1116 IleMetSerGlyAlaGlySerAlaAspGlyGlyTyrSerAspAsnTyrTh 1132

3771 CAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGGA..... 3813
           :::||||||:::      :::|||:::      |||
1132 rHisValGlnValGlyPheAspLysLysHisGluLeuAspGlyValAspL 1149

3814 ................AACACCTTCGACGACGGCATCGGCAACTCGGCA 3846
                       |||:::   ||| :::||||||      ||
1149 euPheThrGlyValThrMetThrTyrThrAsp.SerSerAlaAspSerHi 1165

3847 CGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGA 3896
     |:::              |||||||||:::|||   |||:::|||
1165 sAla.............PheSerGlyLysThrLysSerValGly..Gl 1176

3897 CATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACG 3946
        |||:::   |||:::|||   |||:::|||||||:::      |||
1176 y...GlyLeuTyrAlaSerAlaLeuPheGluSerGlyAlaTyrIleAspL 1192

3947 GCATCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCA 3996
       |||   ||||||              :::|||:::      :::
1192 euIle...GlyLys............TyrIleHisHis......AspAsn 1201

3997 AGATACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGC 4046
        |||   ::::::|||:::|||:::|||
1202 AspTyrThrGlyAsnPheAlaGlyLeuGly..................... 1211

4047 AACGCGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATA 4096
      |||:::::::::         :::   |||
1212 .ThrLysHisTyrAsnThrHisSerTrpTyr..................... 1221

4097 TCGCCACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCA 4146
                                      ||||||   ::::::
1222 ................................AlaGlyAlaGluThr 1226

4147 GATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAG 4196
        |||   :::      :::::::::      |||   |||      ::
1227 GlyTyrArgTyrHisLeuThrGluGluThrPheIleGluProGlnAlaGl 1243

4197 CCTGTCCTATACCGATGCCGCTTCCGGCAAA..................... 4227
     :|||   |||      |||:::||||||||||
1243 uLeuValTyr...GlyAlaValSerGlyLysThrPheArgTrpLysAspG 1259

4228 ................GTCCGAACGCGCGTCAATACCGCCGTATTGGCG 4260
                     ::::::::|||      :::   :::::::::
1259 lyAspMetAspLeuSerMetLysAsnArgAspPheSerProLeuIleGly 1275

4261 CAG.........GATTTCGGCAAAACCCGCAGTGCG...GAATGGGGCGT 4298
     :::         ::::::|||||||||   |||:::   :::|||:::||
1276 ArgThrGlyIleGluLeuGlyLysThrPheSerGlyLysAspTrpSerVa 1292

4299 AAACGCC 4305
     |:::|||
1292 lThrAla 1294
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW27705

seq_documentation_block:

```
ID   AAW27705 standard; Protein; 323 AA.
XX
AC   AAW27705;
XX
DT   08-MAY-1998  (first entry)
XX
DE   H. influenzae Hap protein autotransporter membrane integration region.
XX
KW   Hap protein; autotransporter; Gram-negative bacteria; diagnostic;
KW   therapy; surface presented polypeptide.
XX
OS   Haemophilus influenzae.
XX
PN   WO9735022-A1.
XX
PD   25-SEP-1997.
XX
PF   15-MAR-1996;   96WO-EP01130.
XX
PR   15-MAR-1996;   96WO-EP01130.
XX
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Jose J,  Maurer J,  Meyer TF;
XX
DR   WPI; 1997-480227/44.
DR   N-PSDB; AAT88142.
XX
PT   Presentation of peptide(s) on surface of Gram-negative bacteria -
PT   via transformation with vector encoding signal peptide, presented
PT   peptide and transporter domain of auto-transporter, producing
PT   peptide libraries for epitope mapping
XX
PS   Claim 8; Fig 9; 84pp; German.
XX
CC   This sequence represents the H. influenzae Hap autotransporter membrane
CC   integration region. This region is involved in a novel method which
CC   allows the presentation of stable fusion polypeptides on the surface of
CC   Gram-negative bacteria which can be released into the surrounding media.
CC   The method can be used to produce a variegated population of
CC   surface-presented polypeptides, so that bacteria expressing polypeptides
CC   with particular properties can be identified and simultaneously selected,
CC   e.g. for epitope mapping or selection of ligands with the highest
CC   affinity for antibodies, major histocompatibility complex (MHC) molecules
CC   or other components of the immune system. Selected polypeptides can be
CC   used diagnostically, e.g. to screen sera or antibody banks, and
CC   polypeptide expressing cells may be used as live vaccines. They may also
CC   be used therapeutically, e.g. when the polypeptide is an antibody, to
CC   remove or concentrate pollutants, inactivate toxins, prepare and process
CC   food, prepare washing compositions and label cells. Selected bacteria can
CC   be stored, reproduced and replicated on a large scale as individual
CC   clones.
XX
SQ   Sequence   323 AA;
```

alignment_scores:

```
            Quality:  721.50              Length:    308
              Ratio:  2.898                 Gaps:      3
 Percent Similarity:  80.844   Percent Identity: 45.130
```

alignment_block:

US-09-303-518D-653 x AAW27705 ..

Align seg 1/1 to: AAW27705 from: 1 to: 323

```
3487 CGCCGCGCCCGCCGGGAT...TTGCCGCAACCGCAGCCCCAACCGCAACC 3533
     ::::::::::::::|||   |||   |||   :::   :::   |||
  17 GlnGlnSerGluLysAspArgLeuAlaGlnGluGluAlaGluLysGln.. 32

3534 CCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTG 3583
      :::   |||:::|||||||||||||||||||:::|||||:::|||||||
  33 .ArgLysGlnLysAspLeuIleSerArgTyrSerAsnSerAlaLeuSerG 49

3584 AATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGAC 3633
     ||:::|||||||||:::|||||||:::::::::|||||||||||||||||
  49 luLeuSerAlaThrValAsnSerMetLeuSerValGlnAspGluLeuAsp 65

3634 CGCGTGTTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCG 3683
     |||:::|||:::::::::   ::::::|||||||||||||:::      ::
  66 ArgLeuPheValAspGlnAlaGlnSerAlaValTrpThrAsnIleAlaGl 82

3684 GGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAA. 3732
     :|||   :::::::|||   |||:::   ||||||||||||:::||||||
  82 nAspLysArgArgTyrAspSerAspAlaPheArgAlaTyrGlnGlnGlnL 99
```

```
3733 ..ACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGC 3780
99 ysThrAsnLeuArgGlnIleGlyValGlnLysAlaLeuAlaAsnGlyArg 115

3781 GTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACGG 3830
116 IleGlyAlaValPheSerHisSerArgSerAspAsnThrPheAspGluGl 132

3831 CATCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACG 3880
132 nValLysAsnHisAlaThrLeuThrMetMetSerGlyPheAlaGlnTyrG 149

3881 GCATCGGCAGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGC 3930
149 lnTrpGlyAspLeuGlnPheGlyValAsnValGlyThrGlyIleSerAla 165

3931 GGCAGCCTTTCAGACGGCATCAGAGGCAAAATCCGCCGCCGCCGTGCTGCA 3980
166 SerLysMetAlaGluGluGlnSerArgLysIleHisArgLysAlaIleAs 182

3981 TTACGGCATTCAGGCAAGATACCGCGCAGGTTTCGGCGGATTCGGCATCG 4030
182 nTyrGlyValAsnAlaSerTyrGlnPheArgLeuGlyGlnLeuGlyIleG 199

4031 AACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGA 4080
199 lnProTyrPheGlyValAsnArgTyrPheIleGluArgGluAsnTyrGln 215

4081 TACGAAAACGTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCTACCG 4130
216 SerGluGluValArgValLysThrProSerLeuAlaPheAsnArgTyrAs 232

4131 CGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCA 4180
232 nAlaGlyIleArgValAspTyrThrPheThrProThrAspAsnIleSerV 249

4181 TCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCCGGCAAAGTC 4230
249 alLysProTyrPhePheValAsnTyrValAspValSerAsnAlaAsnVal 265

4231 CGAACGCGCGTCAATACCGCCGTATTGGCGCAGGATTTCGGCAAAACCCG 4280
266 GlnThrThrValAsnLeuThrValLeuGlnGlnProPheGlyArgTyrTr 282

4281 CAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCC 4330
282 pGlnLysGluValGlyLeuLysAlaGluIleLeuHisPheGlnIleSerA 299

4331 TCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCGCAGCACAGCGCG 4380
299 laPheIleSerLysSerGlnGlySerGlnLeuGlyLysGlnGlnAsnVal 315

4381 GGCATCAAATTAGGCTACCGCTGG 4404
316 GlyValLysLeuGlyTyrArgTrp 323
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:AAP80136

seq_documentation_block:

```
ID   AAP80136 standard; protein;  741 AA.
XX
AC   AAP80136;
XX
DT   09-OCT-1990  (first entry)
XX
DE   Neisseria IgA-Protease precursor protein.
XX
KW   IgA-Protease precursor; Neisseria sp;
KW   Gram-negative bacterial live vaccines.
XX
OS   Neisseria gonorrhoeae (strain SMH).
XX
FH   Key             Location/Qualifiers
FT   Cleavage-site   193..197
FT                   /label=IgA-Protease cleavage site (a)
FT   Cleavage-site   226..229
FT                   /label=IgA-Protease cleavage site (b)
FT   Cleavage-site   327..332
FT                   /label=IgA-Protease cleavage site (c)
XX
PN   DE3622221-A.
XX
PD   14-JAN-1988.
XX
PF   02-JUL-1986;   86DE-3622221.
XX
PR   02-JUL-1986;   86DE-3622221.
XX
PA   (PLAC ) MAX PLANCK GES WISSENSCH.
XX
PI   Meyer TF,  Halter R,  Pohlner J;
XX
DR   WPI; 1988-015104/03.
DR   N-PSDB; AAN80154.
XX
PT   Extracellular prodn. of proteins -
PT   by gram-negative host cells contg. a vector contg. one or more
PT   genes coding the desired protein
XX
PS   Disclosure; ; ; German.
XX
CC   Precursor protein consists of three regions i.e. Amino terminal
CC   leader sequence, IgA protease and a "helper" domain. The cleavage
CC   sites given in the features lie in the region between the latter
CC   two domains. DNA encoding a desired protein can be cloned into the
CC   corresponding region in the IgA protease precursor gene, between
CC   the DNA that encodes the natural cleavage sites. Thus, IgA protease
CC   coding region is not disrupted and the desired protein is released
CC   following cleavage by the protease.
XX
SQ   Sequence   741 AA;
```

alignment_scores:

```
           Quality:  442.50              Length:    787
             Ratio:  1.174                 Gaps:     19
 Percent Similarity:  47.903   Percent Identity: 22.490
```

alignment_block:

US-09-303-518D-653 x AAP80136 ..

Align seg 1/1 to: AAP80136 from: 1 to: 741

```
2521 CTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAA 2570
   1 LeuSerAspLysAlaLeuAsnSerPheAspAlaThrProIleAsnGlyAs 17

2571 TGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTA 2620
  17 nValAsnLeuAsnGlnAsnAlaAlaLeuValLeuGlyLysAlaAlaLeuT 34

2621 CCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACACTTA...AAAGAC 2667
  34 rpGlyLysIleGlnGlyGlnGlyAsnSerArgValSerLeuAsnGlnHis 50

2668 AGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAACCTTGA 2717
  51 SerLysTrpHisLeuThrGlyAspSerGlnValHisAsnLeuSerLeuAl 67

2718 CAACGCCACCATTACACTCAATTCCGCCTATCGACACGATGCGGCAGGCG 2767
  67 aAspSerHisIleHisLeuAsnAsnAla......SerAspAlaGlnSerA 82

2768 CGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCC 2817
  82 la..................................................  82
```

```
2818 CTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCT 2867
                                          :::::::::::|||::
  83 ..............................AsnLysTyrHisThrIl 88

2868 GACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGG 2917
     :   :::|||      |||:::|||:::|||   |||:::::::::::::
  88 eLysIleAsn...HisLeuSerGlyAsnGlyHisPheHisTyrLeuThrA 104

2918 AACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAA 2967
     ::|||           :::   |||:::   :::   ||||||::::::
 104 spLeuAlaLysAsnLeuGlyAspLysValLeuValLysGluSerAlaSer 120

2968 GGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCT 3017
     |||   |||   |||   |||:::|||      :::|||||   :::
 121 GlyHisTyrGlnLeuHisValGlnAsnLysThrGlyGluProAsnGln.. 136

3018 CGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCC.... 3063
                      ||||||   |||           |||
 137 ..................GluGlyLeuAspLeuPheAspAlaSerSerV 147

3064 ...........GAAAATCTTAATTTCACCCTGCAAAACGAACACGTCGAT 3102
             :::::::|||      :::|||   |||:::::|||||
 147 alGlnAspArgSerArgLeuPheValSerLeuAlaAsnHisTyrValAsp 163

3103 GCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCT 3152
        ||||||   ||||||   :::      ::::::|||      |||||
 164 LeuGlyAlaLeuArgTyrThrIleLysThrGluAsnGlyIleThrArgLe 180

3153 GCATAAT.....................CCGGTCAAA.............. 3168
     |:::|||                     |||||||||
 180 uTyrAsnProTyrAlaGlyAsnGlyArgProValLysProAlaProSerP 197

3169 ..GAACAAGAGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCC 3216
           :::   :::::::::      ||||||   :::|||::::::
 197 roAlaAlaAsnThrAlaSerGlnAlaGlnLysAlaThrGlnThrAspGly 213

3217 GCCTTGACGGCAAAA................................. 3231
     |||      ||||||
 214 AlaGlnIleAlaLysProGlnAsnIleValValAlaProProSerProGl 230

3232 .......CAGGCACAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACG 3274
            ||||||:::   |||   :::|||||||||:::   :::::::
 230 nAlaAsnGlnAlaGluGluAlaLeuArgGlnGlnAlaLysAlaGluGlnV 247

3275 CG.........CAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCAATGCC 3315
     ::         |||:::   :::|||      :::|||   ::::::
 247 alLysArgGlnGlnAlaAlaGluAlaGluLysValAlaArgGlnLysAsp 263

3316 ACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAG........... 3354
        |||      :::::::::||||||   |||||||||
 264 GluGluAlaLysArgLysAlaAlaGluIleAlaArgGlnGlnGluGluAl 280

3355 ....GCAGGCGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAA 3400
            :::::::|||   |||:::   ::::::||||||   |||:::|
 280 aArgLysAlaAlaGluLeuAlaAlaLysGlnLysAlaGluAlaGluArgL 297

3401 AACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCG 3450
     ||      :::                  ||||||:::|||:::
 297 ysAlaArgGlu.................LeuAlaArgGlnLysAlaGlu 307

3451 GAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCG 3500
     |||:::         ::::::|||   |||:::   ||||||   |||
 308 GluAlaSerHisGlnAlaAsnAlaLysProLysArgArgArgArgArgAl 324

3501 GGATTTGCCGCAACCG................................ 3516
          ||||||:::|||
 324 aIleLeuProArgProProAlaProValPheSerLeuAspAspTyrAspA 341

3516 ..................................................... 3516

 341 laLysAspAsnSerGluSerSerIleGlyAsnLeuAlaArgValIlePro 357
```

```
3516 ..................................................... 3516

 358 ArgMetGlyArgGluLeuIleAsnAspTyrGluGluIleProLeuGluGl 374

3516 ..................................................... 3516

 374 uLeuGluAspGluAlaGluGluGluArgArgGlnAlaThrGlnPheHisS 391

3516 ..................................................... 3516

 391 erLysSerArgAsnArgArgAlaIleSerSerGluProSerSerAspGlu 407

3516 ..................................................... 3516

 408 AspAlaSerGluSerValSerThrSerAspLysHisProGlnAspAsnTh 424

3517 .....................................................C 3517
                                                       |
 424 rGluLeuHisGluLysValGluThrAlaGlyLeuGlnProArgAlaAlaG 441

3518 AGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCC 3567
     ||||||::   |||       |||   |||   :::|||      :::
 441 lnProArgThrGlnAlaAlaAlaGlnAlaAspAlaValSerThrAsnThr 457

3568 AATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTT........ 3609
     ||||||:::|||||||:::      |||:::   :::|||:::
 458 AsnSerAlaLeuSerAspAlaMetAlaSerThrGlnSerIleLeuLeuAs 474

3610 ..........TTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAG 3649
                :::   ::::::         ::::::      |||:::
 474 pThrGlyAlaTyrLeuThrArgHisIleAlaGlnLysSerArgAlaAspA 491

3650 ACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTAC 3699
       ::::::|||:::||||||   |||:::             :::   |||
 491 laGluLysAsnSerValTrpMetSerAsnThrGlyTyrGlyArgAspTyr 507

3700 CGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAAT 3749
        |||   :::::::|||   :::   ::::::            |||||
 508 AlaSerAlaGlnTyrArgArgPheSerSerLysArgThrGlnThrGlnIl 524

3750 CGGTATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTT 3796
     ||||:::::::::::::|||:::::::::   ::::::|||   :::::::
 524 eGlyIleAspArgSerLeuSerGluAsnMetGlnIleGlyGlyValLeuT 541

3797 CGCACAACCGGACCGGAAACACCTTCGACGACGGCATCGGCAACTCGGCA 3846
     ::::::::   :::   :::|||||||||::::::   |||::::::
 541 hrTyrSerAspSerGlnHisThrPheAspGlnAla...GlyGlyLysAsn 556

3847 CGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGC...AGGTT 3893
        ::::::::::::   ::::::|||:::|||   ::::::      ::
 557 ThrPheValGlnAlaAsnLeuTyrGlyLysTyrTyrLeuAsnAspAlaTr 573

3894 CGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAG 3943
     :   ::::::   :::   |||||||||        :::   |||:::
 573 pTyrValAlaGlyAspIleGlyAlaGlySerLeuArgSerArgLeuGlnT 590

3944 ACGGCATCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAG 3993
             :::::::::::::::|||      ::::::   |||:::
 590 hrGlnGlnLysAlaAsnPheAsnArgThrSerIleGlnThrGlyLeuThr 606

3994 GCAAGATACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGG 4043
                     ::::::   |||   |||   |||      ||
 607 LeuGlyAsnThrLeuLysIleAsnGlnPheGluIleValProSerAlaGl 623

4044 CGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCA 4093
     |               :::::::::|||||||||:::      :::   :
 623 yIleArgTyrSerArgLeuSerSerAlaAspTyrLysLeuGlyAspAspS 640

4094 ATATCGCCACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAG 4143
     :::::   :::   :::::::|||   :::         ||||||:::
 640 erValLysValSerSerMetAlaValLysThrLeuThrAlaGlyLeuAsp 656

4144 GCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTT 4193
```

```
            |||   ||||||   :::   ::::::::::::   |||   ||
     657 PheAlaTyrArgPheLys...ValGlyAsnLeuThrValLysProLeuLe 672

    4194 GAGCCTGTCCTATACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCA 4243
         ||||   :::|||      |||      ||||||      :::::::|||:
     672 uSerAlaAlaTyrPhe...AlaAsnTyrGlyLysGlyGlyValAsnValG 688

    4244 ATACCGCCGTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGG 4293
         ::         :::|||            :::   ::::::   :::
     688 lyGlyLysSerPheAlaTyrLysAlaAspAsnGlnGlnGlnTyrSerAla 704

    4294 GGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGC 4343
         ||||||           ::::::   |||||:::::::::::::::
     705 GlyValAlaLeuLeuTyrArgAsnValThrLeuAsnValAsnGlySerIl 721

    4344 CGCCAAGGGGCCGCAATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAG 4393
          :::||||||   |||||||||   |||   |||:::   ||||||:::
     721 eThrLysGlyLysGlnLeuGluLysGlnLysSerGlyGlnIleLysIleG 738

    4394 GCTACCGCTGG 4404
               |||:::
     738 lnIleArgPhe 741
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG98842

seq_documentation_block:

ID AAG98842 standard; Protein; 1569 AA.
XX
AC AAG98842;
XX
DT 26-SEP-2001 (first entry)
XX
DE E. coli growth and proliferation related protein sequence SEQ ID NO:312.
XX
KW Escherichia coli; growth; proliferation; microbial; antimicrobial;
KW bacterial infection; microorganism.
XX
OS Escherichia coli.
XX
PN WO200134810-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30950.
XX
PR 09-NOV-1999; 99US-0164415.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Forsyth RA, Ohlsen K, Zyskind J;
XX
DR WPI; 2001-335933/35.
DR N-PSDB; AAH84513.
XX
PT Novel nucleic acids that inhibit Escherichia coli proliferation, useful
PT for screening for homologous genes and for designing expression vectors
PT -
XX
PS Claim 19; Page 396-399; 522pp; English.
XX
CC AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
CC related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli
CC growth and proliferation related proteins given in AAG99078 and AAG98830
CC to AAG98999. (I) can be used as potential targets for the generation of
CC new antimicrobial agents, and for identification of compounds which
CC interact with the gene products of (I). In addition the expression of
CC (I) and the purification of the proteins, the purified proteins can be
CC used to generate reagents and screen small molecule libraries or other
CC candidate compound libraries for compounds that can be further developed
CC to yield novel antimicrobial compounds. In addition, nucleic acid probes
CC complementary to (I) that are specific for particular species of
CC microorganisms can be used to identify particular microorganism species
CC in clinical specimens, therefore, providing a rapid and dependable
CC method by which to identify the causative agents of a bacterial
CC infection. Also, antibodies generated against proteins translated from
CC mRNA transcribed from proliferation-required sequences can also be used
CC to screen for specific microorganisms that produce such proteins in a
CC species-specific manner. AAH84371 and AAH84670 represent sequencing
CC primers used in the isolation of E. coli growth and proliferation
CC related sequence, which are used in an example from the present
CC invention.
XX
SQ Sequence 1569 AA;

alignment_scores:

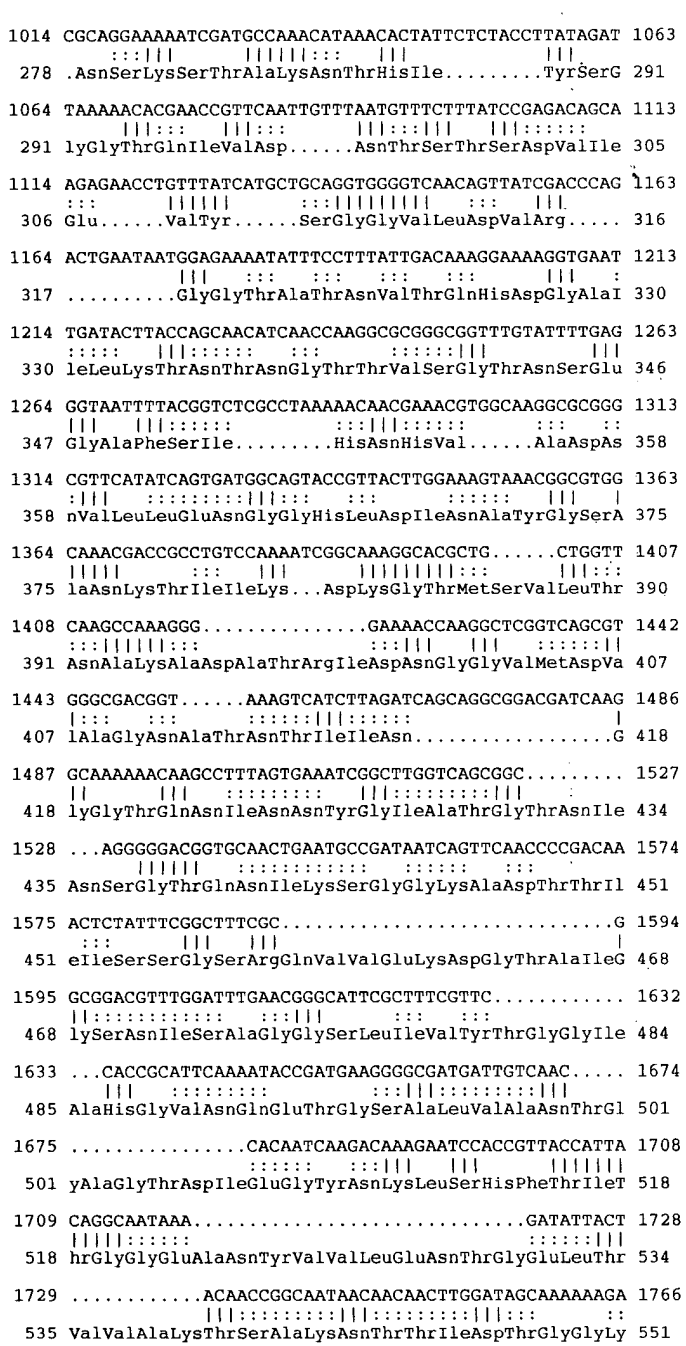
```
            Quality:  359.00              Length:   1686
              Ratio:  0.481                 Gaps:     80
 Percent Similarity:  44.306   Percent Identity: 19.395
```

alignment_block:

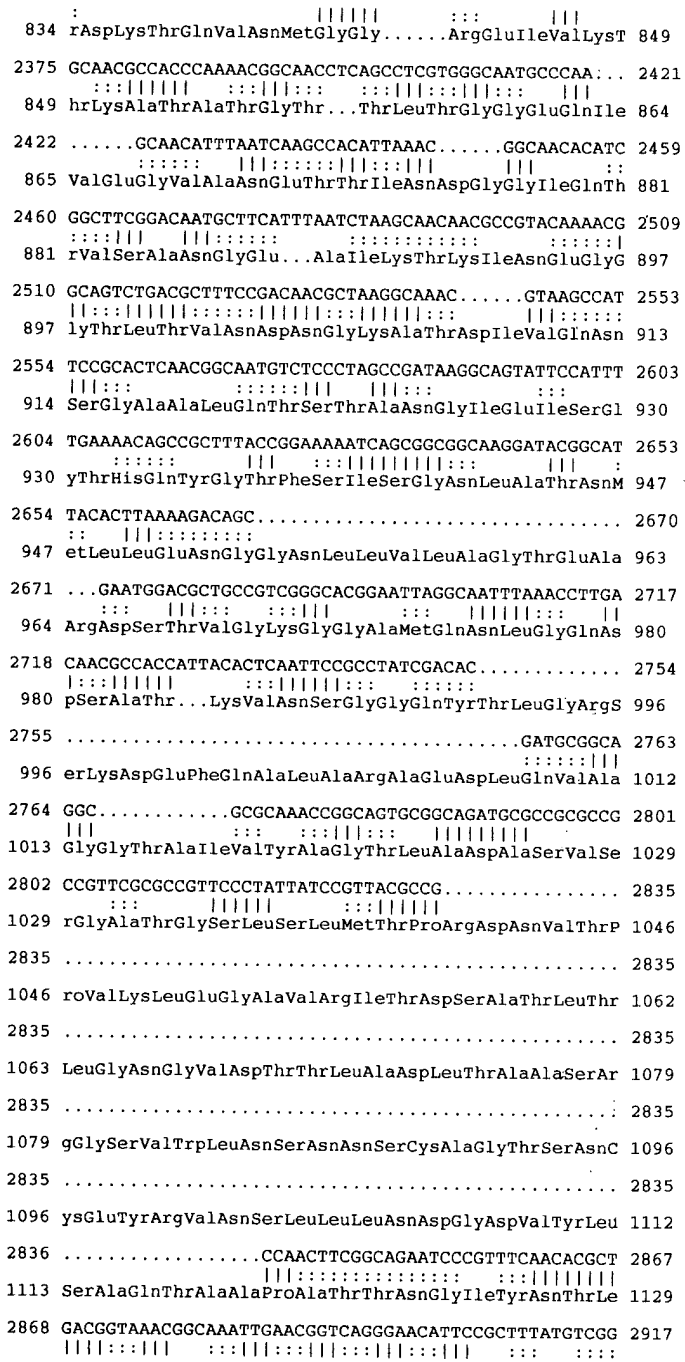
US-09-303-518D-653 x AAG98842 ..

Align seg 1/1 to: AAG98842 from: 1 to: 1569

```
     445 GGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
         ||||||:::::::::   :::      ::::::  :::  ::::::||
      97 GlyThrThrAsnAsnThrHisIleIleAsnHisGlyGlnGlnAsnValHi 113

     495 CAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATG 544
         |      |||:::::::::::::   :::|||      :::|||:::|||
     113 sGlyGlyValSerAsnGlySerLeuIleGluSerGlyGlyTyrGlnAspI 130

     545 GGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATCGGA 594
               ::::::   ::::::           :::::::::   |||:::
     130 leGlySerHisAsnAsnPheValGlyGlnAlaAsnAsnThrThrIleAsn 146

     595 GCAGGCAGACAATATTGGCGGTCTGATGAAGACGAACCCAATAACCGCGA 644
         :::|||||||||
     147 GlyGlyArgGln..................................... 150

     645 AAGTTCATATCATATTGCA......AGCGCATATTCTTGGCTCGTCGGTG 688
             |||   |||   :::      |||:::   :::   :::   :::|
     151 ....SerIleHisAspGlyGlyIleSerThrGlyThrThrIleGluSerG 166

     689 GCAATACCTTTGCACAAAATGGA......TCAGGTGGTGGCACAGTCAAC 732
         |||||      :::   :::|||      |||:::|||   |||::::::
     166 lyAsnGlnAspValTyrLysGlyGlyIleSerAsnGlyThrThrIleLys 182

     733 TTAGGTAGCGAAAAAATTAAA...CATAGCCCATATGGTTTTTTACCAAC 779
            |||:::::::::::::::      |||      |||:::|||
     183 GlyGlyAlaSerArgValGluGlyGlySerAlaAsnGlyIleLeuIleAs 199

     780 AGGAGGCTCA.....................................T 790
          |||||||||                                     :
     199 pGlyGlySerGlnIleValLysValGlnGlyHisAlaAspGlyThrThrI 216

     791 TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
         :::::   |||||||||         :::      :::          :::
     216 leAsnLysSerGlySerGlnAspValValGlnGlySerLeuAlaThrAsn 232

     841 TGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGA..... 885
               |||||||||            |||:::   |||:::
     233 ThrThrIleAsnGly.............GlyArgGlnTyrValGluGlnSe 245

     886 .....................AAAAGCAATGGCTTCCAGCTAGTTCGTA 913
                               |||:::::::|||   |||   |||   :
     245 rThrValGluThrThrThrIleLysAsnGlyGlyGluGlnArgValTyrG 262

     914 AAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTC 963
         :::::          |||      :::   |||   |||:::|||:::
     262 luSerArgAlaLeuAspThrThrIleGluGlyGlyThrGlnSerLeu... 277

     964 TACGAACCACATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGG 1013

     277 ..................................................... 277
```


1014 CGCAGGAAAAATCGATGCCAAACATAAACACTATTCTCTACCTTATAGAT 1063
278 .AsnSerLysSerThrAlaLysAsnThrHisIle.....TyrSerG 291
1064 TAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCA 1113
291 lyGlyThrGlnIleValAsp.....AsnThrSerThrSerAspValIle 305
1114 AGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAG 1163
306 Glu.....ValTyr.....SerGlyGlyValLeuAspValArg..... 316
1164 ACTGAATAATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAAT 1213
317 .....GlyGlyThrAlaThrAsnValThrGlnHisAspGlyAlaI 330
1214 TGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAG 1263
330 leLeuLysThrAsnThrAsnGlyThrThrValSerGlyThrAsnSerGlu 346
1264 GGTAATTTTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGG 1313
347 GlyAlaPheSerIle.....HisAsnHisVal.....AlaAspAs 358
1314 CGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGG 1363
358 nValLeuLeuGluAsnGlyGlyHisLeuAspIleAsnAlaTyrGlySerA 375
1364 CAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTG.....CTGGTT 1407
375 laAsnLysThrIleIleLys...AspLysGlyThrMetSerValLeuThr 390
1408 CAAGCCAAAGGG.....GAAAACCAAGGCTCGGTCAGCGT 1442
391 AsnAlaLysAlaAspAlaThrArgIleAspAsnGlyGlyValMetAspVa 407
1443 GGGCGACGGT.....AAAGTCATCTTAGATCAGCAGGCGGACGATCAAG 1486
407 lAlaGlyAsnAlaThrAsnThrIleIleAsn.....G 418
1487 GCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGC..... 1527
418 lyGlyThrGlnAsnIleAsnAsnTyrGlyIleAlaThrGlyThrAsnIle 434
1528 ...AGGGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAA 1574
435 AsnSerGlyThrGlnAsnIleLysSerGlyGlyLysAlaAspThrThrIl 451
1575 ACTCTATTTCGGCTTTCGC.....G 1594
451 eIleSerSerGlySerArgGlnValValGluLysAspGlyThrAlaIleG 468
1595 GCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTC..... 1632
468 lySerAsnIleSerAlaGlyGlySerLeuIleValTyrThrGlyGlyIle 484
1633 ...CACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAAC..... 1674
485 AlaHisGlyValAsnGlnGluThrGlySerAlaLeuValAlaAsnThrGl 501
1675 .....CACAATCAAGACAAAGAATCCACCGTTACCATTA 1708
501 yAlaGlyThrAspIleGluGlyTyrAsnLysLeuSerHisPheThrIleT 518
1709 CAGGCAATAAA.....GATATTACT 1728
518 hrGlyGlyGluAlaAsnTyrValValLeuGluAsnThrGlyGluLeuThr 534
1729 .....ACAACCGGCAATAACAACAACTTGGATAGCAAAAAAGA 1766
535 ValValAlaLysThrSerAlaLysAsnThrThrIleAspThrGlyGlyLy 551


1767 AATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAACG 1816
551 sLeuIleVal.....GlnLysGluAlaLysThrAspSerT 563
1817 GGCGGCTCAAT.....CTG 1830
563 hrArgLeuAsnAsnGlyGlyValLeuGluValGlnAspGlyGlyGluAla 579
1831 AATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCC..... 1872
580 LysHisValGluGlnGlnSerGlyGlyAlaLeuIleAlaSerThrThrSe 596
1873 .....GGCGGAACAAATTTAAACGGCAAT..... 1896
596 rGlyThrLeuIleGluGlyThrAsnSerTyrGlyAspAlaPheTyrIleA 613
1896 1896
613 rgAsnSerGluAlaLysAsnValValLeuGluAsnAlaGlySerLeuThr 629
1897 .....ATCACGCAAACAAACGGCAA 1916
630 ValValThrGlySerArgAlaValAspThrIleIleAsnAlaAsnGlyLy 646
1917 ACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAA 1966
646 sMetAspValTyrGlyLysAspValGlyThrValLeuAsnSerAlaGlyT 663
1967 GC..... 1968
663 hrGlnThrIleTyrAlaSerAlaThrSerAspLysAlaAsnIleLysGly 679
1969 .....GGGTGGTCAAAAATGGAAGGTATCCCACAAGG 2000
680 GlyLysGlnThrValTyrGlyLeuAlaThrGluAlaAsnIleGluSerGl 696
2001 AGAAATCGTGTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGAAA 2050
696 yGluGlnIleValAspGlyGlySerThrGluLysThr..... 708
2051 ACTTCCATATTCAGGGCGGA..... 2070
709 .....HisIleAsnGlyGlyThrGlnThrValGlnAsnTyrGlyLysAla 723
2071 .....CAAGCGGTGGTTTCCCGCAATGT 2093
724 IleAsnThrAspIleValSerGlyLeuGlnGlnIleMetAlaAsnGlyTh 740
2094 TGCCAAA.....GTGGAAGGCGATTGGCATTTAAGCAATCACGCCC 2134
740 rAlaGluGlySerIleIleAsnGlyGlySerGlnValValAsnGluGlyG 757
2135 AAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGT 2184
757 lyLeuAlaGluAsnSerValLeuAsnAspGlyGlyThrLeuAspValArg 773
2185 .....TCGGACTGGACGGGTCTG..... 2202
774 GluLysGlySerAlaThrGlyIleGlnGlnSerSerGlnGlyAlaLeuVa 790
2203 .....ACAAGTTGTACCGAAAAAACCATTACCGACGATAAAGTGATTG 2245
790 lAlaThrThrArgAlaThrArgValThrGlyThrArgAlaAspGlyValA 807
2246 CTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGATCAC 2295
807 laPheSerIleGluGlnGlyAlaAlaAsnAsnIleLeuLeuAlaAsn... 822
2296 GCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGC 2345
823 .....GlyGlyValLeuThrValGluSerAspThrSerSe 834
2346 A.....GGCGGAGACACGCACTATACGGTTACGC 2374


834 rAspLysThrGlnValAsnMetGlyGly.....ArgGluIleValLysT 849
2375 GCAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAA... 2421
849 hrLysAlaThrAlaThrGlyThr...ThrLeuThrGlyGlyGluGlnIle 864
2422 .....GCAACATTTAATCAAGCCACATTAAAC.....GGCAACACATC 2459
865 ValGluGlyValAlaAsnGluThrThrIleAsnAspGlyGlyIleGlnTh 881
2460 GGCTTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACG 2509
881 rValSerAlaAsnGlyGlu...AlaIleLysThrLysIleAsnGluGlyG 897
2510 GCAGTCTGACGCTTTCCGACAACGCTAAGGCAAAC.....GTAAGCCAT 2553
897 lyThrLeuThrValAsnAspAsnGlyLysAlaThrAspIleValGlnAsn 913
2554 TCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTT 2603
914 SerGlyAlaAlaLeuGlnThrSerThrAlaAsnGlyIleGluIleSerGl 930
2604 TGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCAT 2653
930 yThrHisGlnTyrGlyThrPheSerIleSerGlyAsnLeuAlaThrAsnM 947
2654 TACACTTAAAAGACAGC..... 2670
947 etLeuLeuGluAsnGlyGlyAsnLeuLeuValLeuAlaGlyThrGluAla 963
2671 ...GAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAACCTTGA 2717
964 ArgAspSerThrValGlyLysGlyGlyAlaMetGlnAsnLeuGlyGlnAs 980
2718 CAACGCCACCATTACACTCAATTCCGCCTATCGACAC..... 2754
980 pSerAlaThr...LysValAsnSerGlyGlyGlnTyrThrLeuGlyArgS 996
2755 .....GATGCGGCA 2763
996 erLysAspGluPheGlnAlaLeuAlaArgAlaGluAspLeuGlnValAla 1012
2764 GGC.....GCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCG 2801
1013 GlyGlyThrAlaIleValTyrAlaGlyThrLeuAlaAspAlaSerValSe 1029
2802 CCGTTCGCGCCGTTCCCTATTATCCGTTACGCCG..... 2835
1029 rGlyAlaThrGlySerLeuSerLeuMetThrProArgAspAsnValThrP 1046
2835 2835
1046 roValLysLeuGluGlyAlaValArgIleThrAspSerAlaThrLeuThr 1062
2835 2835
1063 LeuGlyAsnGlyValAspThrThrLeuAlaAspLeuThrAlaAlaSerAr 1079
2835 2835
1079 gGlySerValTrpLeuAsnSerAsnAsnSerCysAlaGlyThrSerAsnC 1096
2835 2835
1096 ysGluTyrArgValAsnSerLeuLeuLeuAsnAspGlyAspValTyrLeu 1112
2836 .....CCAACTTCGGCAGAATCCCGTTTCAACACGCT 2867
1113 SerAlaGlnThrAlaAlaProAlaThrThrAsnGlyIleTyrAsnThrLe 1129
2868 GACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGG 2917
1129 uThrThrAsn...GluLeuSerGlySerGlyAsnPheTyrLeuHisThrA 1145
2918 AACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAA 2967
1145 snValAlaGlySerArgGlyAspGlnLeuValValAsnAsnAsnAlaThr 1161
2968 GGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCT 3017
1162 GlyAsnPheLysIlePheValGlnAspThrGlyValSerProGlnSerAs 1178
3018 CGAGCAATTGACGGTAGTGGAA......GGAAAAGACAACACACCGCTGT 3061
1178 pAspAlaMetThrLeuValLysThrGlyGlyGlyAsp..... 1190
3062 CCGAAAATCTTAATTTCACCCTGCAAAACGAA......CACGTCGATGCC 3105
1191 .....AlaSerPheThrLeuGlyAsnThrGlyGlyPheValAspLeu 1204
3106 GGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCA 3155
1205 GlyThrTyrGluTyr...ValLeuLysSerAspGly.AsnSer..... 1217
3156 TAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCGGGAG 3205
1217 1217
3206 AAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCCAAACAA 3255
1217 1217
3256 CAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGG 3305
1218 .....AsnTrpAsnLeuThrAsnAspValLysProA 1228
3306 GCGCAATGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGG 3355
1228 snProAspPro.....IleProAsn..... 1234
3356 CAGGCGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGG 3405
1234 1234
3406 GTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAAC 3455
1235 .....ProLysP 1237
3456 CCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATT 3505
1237 roAspProLysProAspProLysProAspPro..AsnProLysProAsp. 1252
3506 TGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATC 3555
1253 ..ProThrProAspProThrProThrProValProGluLysArgIleThr 1268
3556 AGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAG 3605
1269 Pro.....SerThrAlaAlaValLeuAsnMetAlaAlaThrLeuProLe 1283
3606 CGTTTTC.....GCCGTACAGGACGAATTGGACCGCGTGT 3640
1283 uValPheAspAlaGluLeuAsnSerIleArgGluArgLeuAsnIleMetL 1300
3641 TTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACC 3690
1300 ysAlaSerProHisAsnAsnAsnValTrp...GlyAlaThrTyrAsnThr 1315
3691 AAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCT 3740
1316 ArgAsnAsnValThrThrAspAlaGlyAlaGlyPheGluGlnThr..... 1330
3741 GCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGC.....G 3781
1331 ....LeuThrGlyMetThrValGlyIleAspSerArgAsnAspIleProG 1346

```
3782 TCGGCATCCTG.....................TTTTCGCACAACCGGACC 3810
       ||||||                        :::||||||::::::
1346 luGlyIleThrThrLeuGlyAlaPheMetGlyTyrSerHisSerHisIle 1362

3811 GGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGG 3860
     |||      ||||||   |||                    :::||||||
1363 Gly......PheAspArgGly.....................GlyHisGl 1370

3861 TGCCGTTTTCGGGCAATACGGCATCGGC..................... 3888
     |:::|||   |||:::|||:::::|||
1370 ySerVal...GlySerTyrSerLeuGlyGlyTyrAlaSerTrpGluHisG 1386

3889 ....................................AGGTTCGACATC 3900
                                         ||||||
1386 luSerGlyPheTyrLeuAspGlyValValLysLeuAsnArgPheLysSer 1402

3901 GGCATCAGC.........GCGGGCGCGGGTTTTAGTAGCGGCAGCCTT.. 3939
     :::::::::        ::::::::::|||   ::::::||||||
1403 AsnValAlaGlyLysMetSerSerGlyGlyAlaAlaAsnGlySerTyrHi 1419

3940 .TCAGACGGCATCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCA 3988
      |||:::|||:::   |||                            |
1419 sSerAsnGlyLeuGlyGlyHis.........................I 1427

3989 TTCAGGCAAGATACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCAC 4038
     ||::::::      |||         |||:::::::::::::   |||:::
1427 leGluThrGlyMetArgPheThrAspGlyAsnTrpAsnLeuThrProTyr 1443

4039 ATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAA 4088
        :::   |||   :::   :::::::::   :::|||:::   :::||
1444 AlaSerLeuThrGlyPheThrAlaAspAsnProGluTyrHisLeuSerAs 1460

4089 C............GTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCT 4126
     |            ::::::   |||   ::::::            |
1460 nGlyMetLysSerLysSerValAspThrArgSerIle...........T 1473

4127 ACCGC...GCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACAC 4173
     |||||      |||         :::|||::::::::::   ::::::
1473 yrArgGluLeuGlyAlaThrLeuSerTyrAsnMetArgLeuGlyAsnGly 1489

4174 ATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCCGG 4223
     :::::::::   |||:::|||:::   :::
1490 MetGluValGluProTrpLeuLysAlaAlaValArgLysGluPheValAs 1506

4224 CAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCGCAGGATTTC.... 4269
            |||::::::||||||:::              ::::::
1506 pAspAsnArgValLysValAsnSerAspGlyAsnPheValAsnTyrLeuS 1523

4270 ..GGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGT 4317
       |||:::           :::   |||::::::|||::::::::::::
1523 erGlyArgArgGlyIleTyrGlnAlaGlyIleLysAlaSerPheSerSer 1539

4318 TTCACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGC 4367
       |||||||||   |||   :::::::::
1540 ...ThrLeuSerGlyHisLeuGlyValGly.................Ty 1549

4368 GCAGCACAGCGCGGGCATCAAA.................TTAGGCTACC 4399
      :::||||||||||||::::::                 |||   :
1549 rSerHisSerAlaGlyValGluSerProTrpAsnAlaValAlaGlyValA 1566

4400 GCTGG 4404
     ::|||
1566 snTrp 1567
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01830

seq_documentation_block:
ID AAB01830 standard; Protein; 1222 AA.
XX
AC AAB01830;
XX
DT 11-SEP-2000 (first entry)
XX
DE H. influenzae strain K1 mature full-length HMW1A protein, SEQ ID NO:37.
XX
KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
OS Haemophilus influenzae strain K1.
XX
FH Key Location/Qualifiers
FT Misc-difference 307
FT /note= "Encoded by GG"
XX
PN WO200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
DR WPI; 2000-303789/26.
DR N-PSDB; AAA52180.
XX
PT Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
PS Claim 8; Fig 20A-R; 307pp; English.
XX
CC The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMWA
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typeable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents a mature HMWA protein from a non-typeable
CC strain of H. influenzae.
XX
SQ Sequence 1222 AA;

alignment_scores:
Quality: 310.50 Length: 1304
Ratio: 0.475 Gaps: 62
Percent Similarity: 50.153 Percent Identity: 19.785

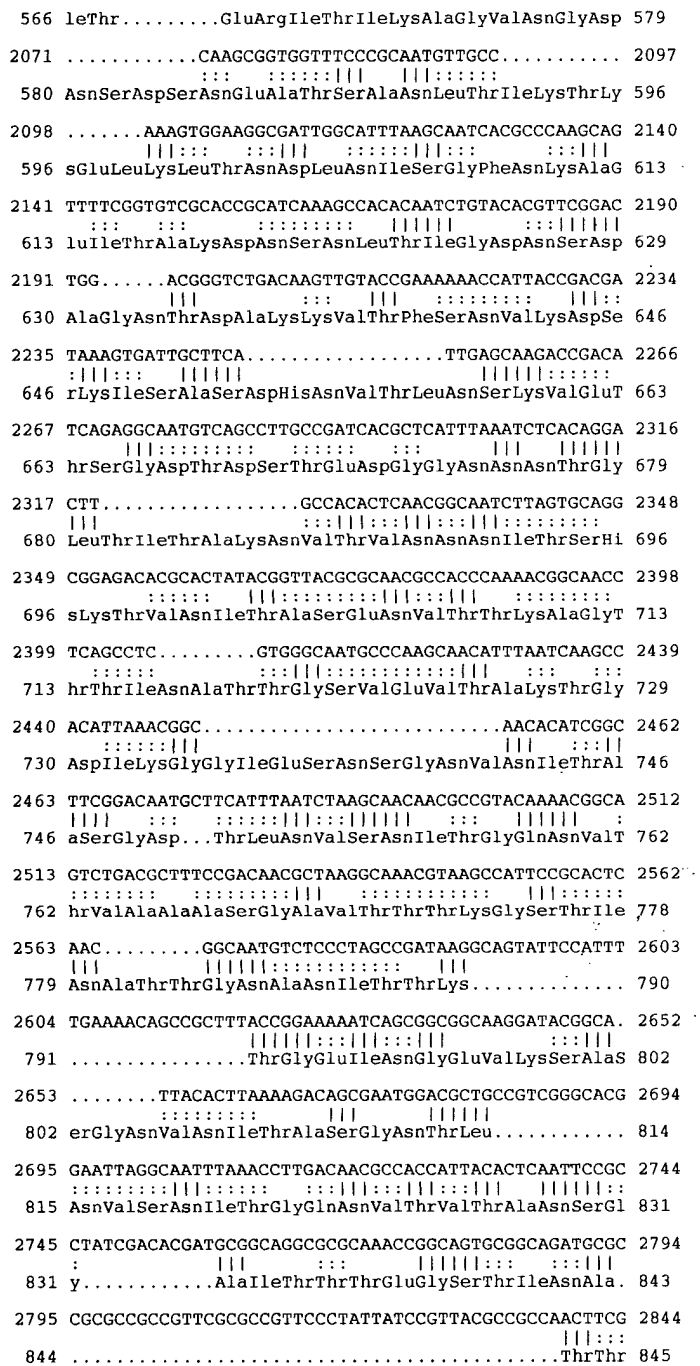
alignment_block:

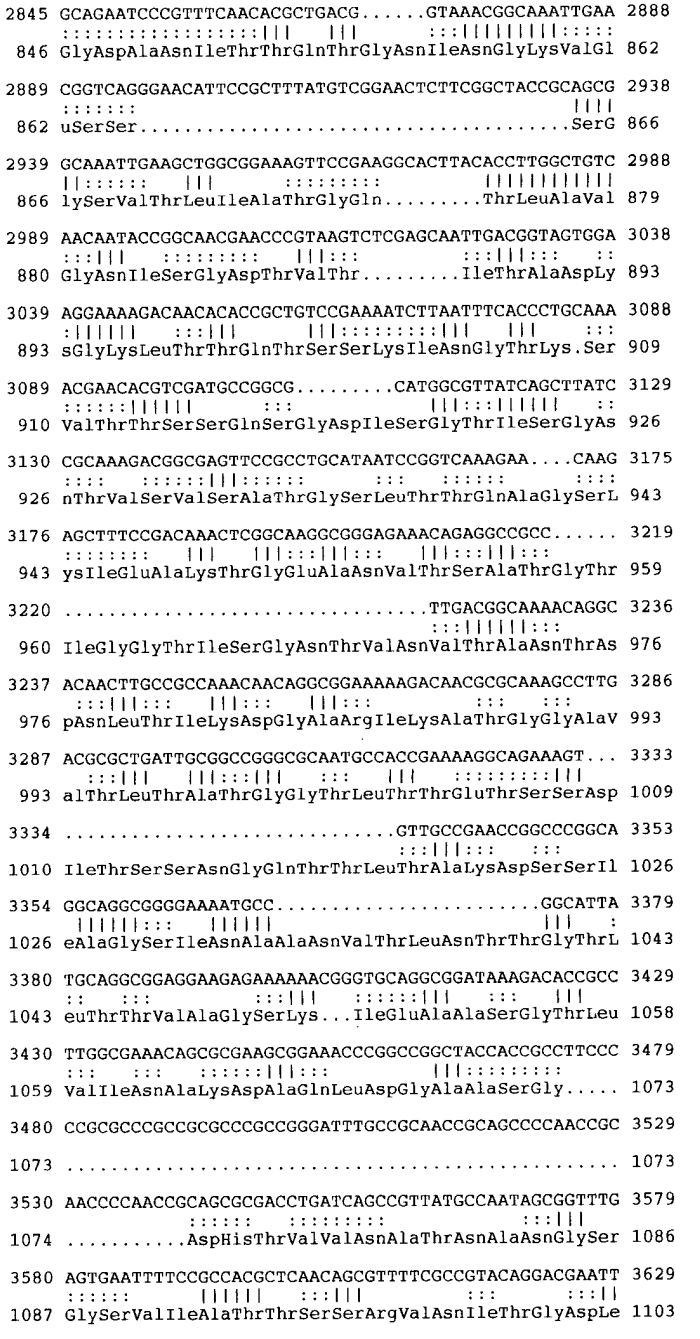
US-09-303-518D-653 x AAB01830 ..

Align seg 1/1 to: AAB01830 from: 1 to: 1222

```
 502 GTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATGGGTGGAA 551
     :::|||:::::          :::::::|||   |||      ||||||
  95 IleThrSerThrThrAsnGlyAsnLeuThrIleTyrSerGlyGlyTrp.. 110

 552 ATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATCGGAGCAGGC. 600
       :::|||:::::::|||         ::::::   :::   :::|||
 111 ....ValAspValHisLys.........AsnIleThrLeuLysSerGlyT 123

 601 ................AGACAATATTGGCGGTCTGATGAAGACGAACCC 633
                     :::|||         :::   ||||||:::|||
 123 yrLeuAsnIleThrThrLysGlnGlyAspIleAlaPheGluAspLysPro 139

 634 AAT.....................................AACCG 641
     :::                                    |||::
 140 GlyLeuSerAsnLeuThrIleThrAlaLysGlyThrIleAlaValAsnAs 156

 642 CGAAAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCA 691
     ::::::::::::::::::                     . :
 156 nLysLysGlyPheArgPhe.........................A 163

 692 ATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGC 741
     :::::   :::   ||||||:::||||||||||
 163 spAsnValThrLeuAsnGlyThrGlyGlyGlyLeu.............. 174

 742 GAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGA........ 783
                         :::   ::::::   ||||||
 175 ..................SerPheLysTyrIleGluThrGlyAsnArgAs 185

 784 .........................GGCTCATTTGGCGACAGTGGCTCAC 808
                             |||   ::::::   ||||||:::
 185 pSerAsnPheGluThrHisPheArgGlyArgLeuAsnIleSerGlyLysV 202

 809 CAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTA 858
       ::::::   :::|||:::::::::::|||
 202 alAspIleLeuMetGlnAlaArgGlnGluAsnTrp.............. 213

 859 TTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTAGT 908

 214 ..............................................As 214

 909 TCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAG 958
      |||:::   |||                  |||   :::|||
 214 nArgArgHisTrp.....................GlyArgSerHisTrpA 224

 959 TATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAAC......... 999
             :::   :::   :::   :::   :::||||||
 224 snValThrArgLeuAsnValSerGluAsnSerTyrPheAsnValThrIle 240

1000 ..............GACAATAATAATGGCGCAGGAAAAATCGATGCCAA 1034
                    ::::::   |||||||||   :::::::|||::
 241 AspSerSerGlySerAlaSerSerProGlyAlaGlyProLeuAsnAlaGl 257

1035 ACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAAT 1084
     :          ::::::   ::::::      |||           :
 257 nSerGlyLeuAsnGlyIleSerPheAsnAsnAspThr............V 270

1085 TGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCT 1134
     ::||||||::::::                                |||
 270 alPheAsnIleAla..............................Ala 275

1135 GCAGGTGGGGTCAAC...AGTTATCGACCC............AGACTGAA 1169
     :::::::::||||||   :::   :::|||          ::::::::
 276 SerSerAlaValAsnPheAsnIleLysProProIleValAspLysValTh 292

1170 TAATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATAC 1219
     :||||||::::::      |||            ||||||::::::   :
 292 rAsnGlyAsnHisThrLeuPhe............LysGlyAsnIleSerV 305

1220 TTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAAT 1269
```

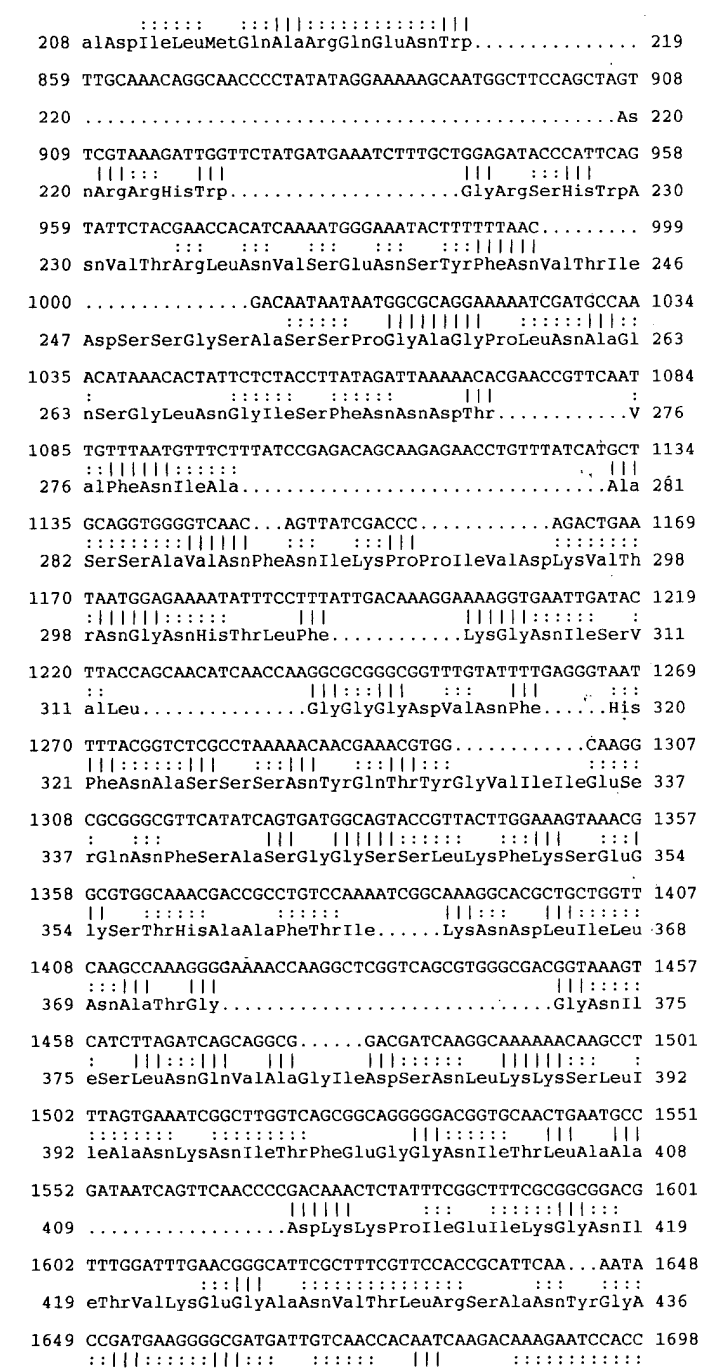
```
      ::                  |||:::|||   :::   |||      :::
  305 alLeu..............GlyGlyGlyAspValAsnPhe......His 314

 1270 TTTACGGTCTCGCCTAAAAACAACGAAACGTGG............CAAGG 1307
      |||::::::|||   :::|||   :::|||:::           :::::
  315 PheAsnAlaSerSerSerAsnTyrGlnThrTyrGlyValIleIleGluSe 331

 1308 CGCGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACG 1357
      :   :::          |||   ||||||::::::   :::|||   :::|
  331 rGlnAsnPheSerAlaSerGlyGlySerSerLeuLysPheLysSerGluG 348

 1358 GCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTT 1407
      ||   ::::::      ::::::          |||:::   |||:::::::
  348 lySerThrHisAlaAlaPheThrIle......LysAsnAspLeuIleLeu 362

 1408 CAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGT 1457
      :::|||   |||                              |||:::::
  363 AsnAlaThrGly..............................GlyAsnIl 369

 1458 CATCTTAGATCAGCAGGCG......GACGATCAAGGCAAAAAACAAGCCT 1501
      :   |||:::|||   |||      |||::::::   |||||||:::   :
  369 eSerLeuAsnGlnValAlaGlyIleAspSerAsnLeuLysLysSerLeuI 386

 1502 TTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCC 1551
      ::::::::   :::::::::         |||::::::   |||   |||
  386 leAlaAsnLysAsnIleThrPheGluGlyGlyAsnIleThrLeuAlaAla 402

 1552 GATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACG 1601
                      ||||||      :::   ::::::|||:::
  403 .................AspLysLysProIleGluIleLysGlyAsnIl 413

 1602 TTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAA...AATA 1648
                ::::||   ::::::::::::::      :::   ::::
  413 eThrValLysGluGlyAlaAsnValThrLeuArgSerAlaAsnTyrGlyA 430

 1649 CCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACC 1698
      ::|||:::::::|||:::   ::::::   |||      ::::::::::::
  430 snAspLysSerAlaLeuSerIleArgGlyAsnValThrAsnLysGlyAsn 446

 1699 GTTACCATTACAGGCAAT..............AAAGATATTACTACAAC 1733
      :::|||:::||||||:::              |||:::::|||:::
  447 LeuThrValThrGlySerAlaIleAsnIleGluLysAsnLeuThrValGl 463

 1734 CGGCAAT..................................... 1740
       |||:::
  463 uGlySerAlaLysPheLeuAlaAsnProAsnTyrSerPheAsnValSerG 480

 1741 ........AACAACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGT 1782
                :::|||:::      |||:::   :::||||||   :::|||
  480 lyLeuPheAspAsnGlnGlyLysSerAsnIleSerIleAlaLysGlyGly 496

 1783 TGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAA 1832
         :::      ||||||      :::|||   |||   ||||||::::
  497 AlaIle...PheLysAspIleGluAsnThr...GlySerLeuAsnIleTh 511

 1833 TTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAA 1882
      :   :::   :::::::      :::|||::::::   ::::::   ||||
  511 rThrLysSerAspSerAsnHisHisThrIleIleLysGlyAsnIleThrA 528

 1883 ATTTAAACGGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGC 1932
      ||   :::|||::::::::::   ||||||               :::|||
  528 snArgLysGlyAspLeuAsnIleThrAsn..............AsnGly 539

 1933 AGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAAT 1982
            |||               :::::::|||::::::   |||:::
  540 AspAsnThrGlu.........IleGlnIleGlyGlyAsnIleSerGlnLy 553

 1983 GGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACC 2032
       |||            |||:::::::::   ::::::|||   ::::::
  553 sGlu............GlyAsnLeuThrIleSerSerAspLysValAsnI 566

 2033 GCACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGA............ 2070
       |||         |||::::::   |||:::::::|||
```

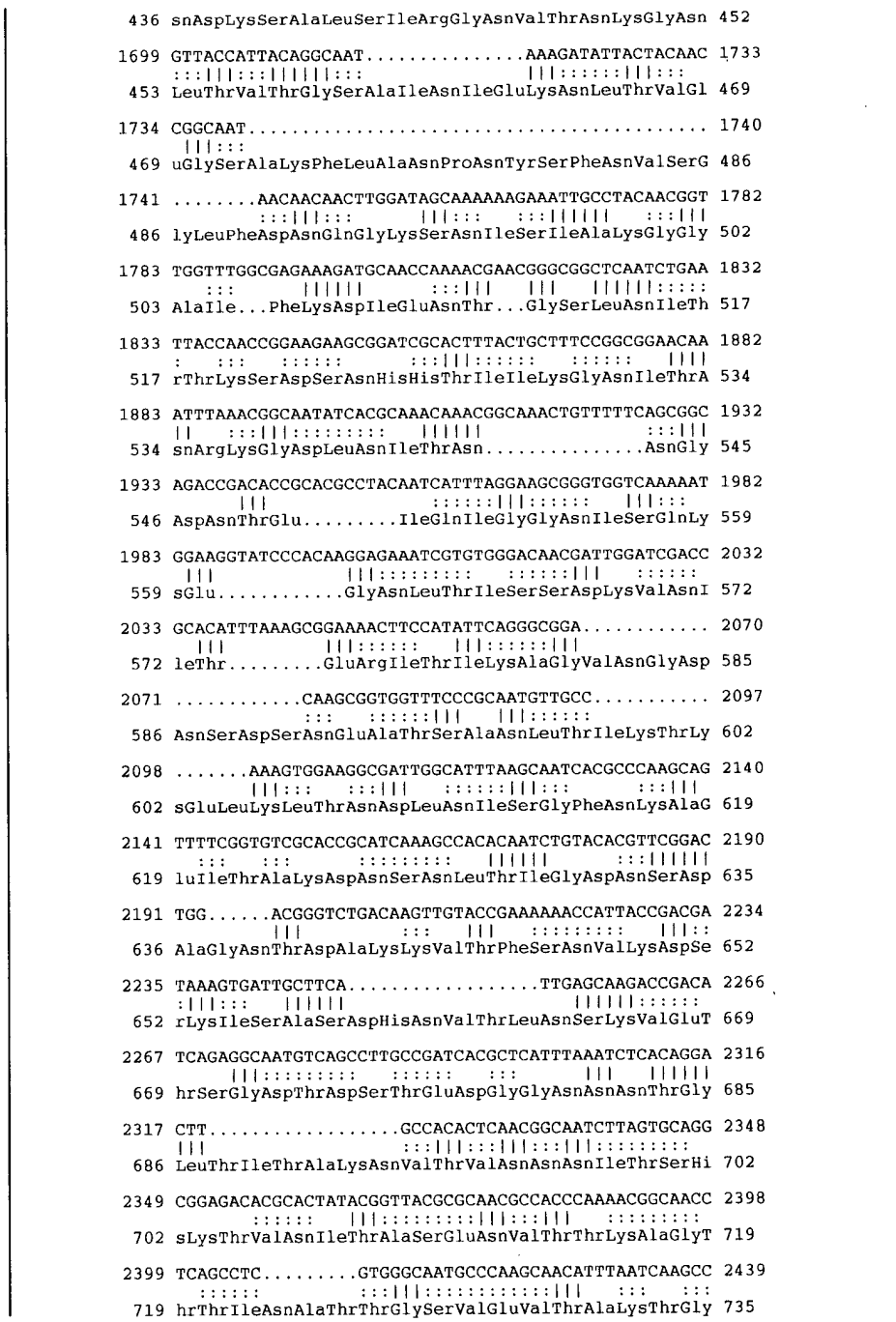

566 leThr.....GluArgIleThrIleLysAlaGlyValAsnGlyAsp 579
2071 .....CAAGCGGTGGTTTCCCGCAATGTTGCC..... 2097
580 AsnSerAspSerAsnGluAlaThrSerAlaAsnLeuThrIleLysThrLy 596
2098 .....AAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCAG 2140
596 sGluLeuLysLeuThrAsnAspLeuAsnIleSerGlyPheAsnLysAlaG 613
2141 TTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGAC 2190
613 luIleThrAlaLysAspAsnSerAsnLeuThrIleGlyAspAsnSerAsp 629
2191 TGG.....ACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGACGA 2234
630 AlaGlyAsnThrAspAlaLysLysValThrPheSerAsnValLysAspSe 646
2235 TAAAGTGATTGCTTCA.....TTGAGCAAGACCGACA 2266
646 rLysIleSerAlaSerAspHisAsnValThrLeuAsnSerLysValGluT 663
2267 TCAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGA 2316
663 hrSerGlyAspThrAspSerThrGluAspGlyGlyAsnAsnAsnThrGly 679
2317 CTT.....GCCACACTCAACGGCAATCTTAGTGCAGG 2348
680 LeuThrIleThrAlaLysAsnValThrValAsnAsnAsnIleThrSerHi 696
2349 CGGAGACACGCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACC 2398
696 sLysThrValAsnIleThrAlaSerGluAsnValThrThrLysAlaGlyT 713
2399 TCAGCCTC.....GTGGGCAATGCCCAAGCAACATTTAATCAAGCC 2439
713 hrThrIleAsnAlaThrThrGlySerValGluValThrAlaLysThrGly 729
2440 ACATTAAACGGC.....AACACATCGGC 2462
730 AspIleLysGlyGlyIleGluSerAsnSerGlyAsnValAsnIleThrAl 746
2463 TTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGCA 2512
746 aSerGlyAsp...ThrLeuAsnValSerAsnIleThrGlyGlnAsnValT 762
2513 GTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTC 2562
762 hrValAlaAlaAlaSerGlyAlaValThrThrThrLysGlySerThrIle 778
2563 AAC.....GGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTT 2603
779 AsnAlaThrThrGlyAsnAlaAsnIleThrThrLys..... 790
2604 TGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCA. 2652
791 .....ThrGlyGluIleAsnGlyGluValLysSerAlaS 802
2653 .....TTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACG 2694
802 erGlyAsnValAsnIleThrAlaSerGlyAsnThrLeu..... 814
2695 GAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGC 2744
815 AsnValSerAsnIleThrGlyGlnAsnValThrValThrAlaAsnSerGl 831
2745 CTATCGACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGC 2794
831 y.....AlaIleThrThrThrGluGlySerThrIleAsnAla. 843
2795 CGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCG 2844
844 .....ThrThr 845


2845 GCAGAATCCCGTTTCAACACGCTGACG.....GTAAACGGCAAATTGAA 2888
846 GlyAspAlaAsnIleThrThrGlnThrGlyAsnIleAsnGlyLysValGl 862
2889 CGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCG 2938
862 uSerSer.....SerG 866
2939 GCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTC 2988
866 lySerValThrLeuIleAlaThrGlyGln.....ThrLeuAlaVal 879
2989 AACAATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGGA 3038
880 GlyAsnIleSerGlyAspThrValThr.....IleThrAlaAspLy 893
3039 AGGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAA 3088
893 sGlyLysLeuThrThrGlnThrSerSerLysIleAsnGlyThrLys.Ser 909
3089 ACGAACACGTCGATGCCGGCG.....CATGGCGTTATCAGCTTATC 3129
910 ValThrThrSerSerGlnSerGlyAspIleSerGlyThrIleSerGlyAs 926
3130 CGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAA....CAAG 3175
926 nThrValSerValSerAlaThrGlySerLeuThrThrGlnAlaGlySerL 943
3176 AGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCC..... 3219
943 ysIleGluAlaLysThrGlyGluAlaAsnValThrSerAlaThrGlyThr 959
3220 .....TTGACGGCAAAACAGGC 3236
960 IleGlyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnThrAs 976
3237 ACAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTG 3286
976 pAsnLeuThrIleLysAspGlyAlaArgIleLysAlaThrGlyGlyAlaV 993
3287 ACGCGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGT... 3333
993 alThrLeuThrAlaThrGlyGlyThrLeuThrThrGluThrSerSerAsp 1009
3334 .....GTTGCCGAACCGGCCCGGCA 3353
1010 IleThrSerSerAsnGlyGlnThrThrLeuThrAlaLysAspSerSerIl 1026
3354 GGCAGGCGGGGAAAATGCC.....GGCATTA 3379
1026 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrL 1043
3380 TGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCC 3429
1043 euThrThrValAlaGlySerLys...IleGluAlaAlaSerGlyThrLeu 1058
3430 TTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCC 3479
1059 ValIleAsnAlaLysAspAlaGlnLeuAspGlyAlaAlaSerGly..... 1073
3480 CCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGC 3529
1073 1073
3530 AACCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTG 3579
1074 .....AspHisThrValValAsnAlaThrAsnAlaAsnGlySer 1086
3580 AGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATT 3629
1087 GlySerValIleAlaThrThrSerSerArgValAsnIleThrGlyAspLe 1103

```
     3630 GGAC..............CGCGTGTTTGCCGAAGACCGCCGCAACGCCG 3664
          |                 ::::::::::::::::::   :::|||:::|
     1103 uIleThrIleAsnGlyLeuAsnIleIleSerLysAsnGlyLysAsnThrV 1120

     3665 TTTGGACAAGCGGCATCCGG...GACACCAAACACTACCGTTCG..... 3705
          ||      :::|||::::::   |||:::|||:::   :::
     1120 alLeuLeuLysGlyValGluIleAspValLysTyrIleGlnProGlyIle 1136

     3706 ....................CAAGATTTCCGCGCCTACCGCCAACAAAC 3734
                              :::     ||||||   ::::::
     1137 AlaSerValAsnGluValIleGluAlaLysArgAlaLeuGluLysValLy 1153

     3735 CGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCG 3784
           |||||| :::   :::      :::      :::      |||:::|||
     1153 sAspLeu.SerAspGluGluArgGluThrLeuAlaLysLeuGlyValSer 1169

     3785 GCATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACGGCATC 3834
          |||      |||               |||      ::::::   :::
     1170 AlaValArgPhe............AlaGluProAsnAsnAlaIleThrIl 1182

     3835 GGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCAT 3884
            |||::::::   :::   |||   |||:::|||      |||   |
     1182 eAsnThrGlnAsnGluPheThrThrArgProLeuSerGlnValThrIleS 1199

     3885 CGGCAGGT 3892
          ||   |||
     1199 erGluGly 1201
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01828

seq_documentation_block:
ID   AAB01828 standard; Protein; 1228 AA.
XX
AC   AAB01828;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain K1 full-length HMW1A protein, SEQ ID NO:34.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain K1.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 313
FT                   /note= "Encoded by GG"
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52179.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 20A-R; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1228 AA;


alignment_scores:
            Quality:  310.50              Length:   1304
              Ratio:  0.475                Gaps:     62
 Percent Similarity:  50.153   Percent Identity: 19.785

alignment_block:
 US-09-303-518D-653 x AAB01828   ..

 Align seg 1/1  to: AAB01828  from: 1  to: 1228

     502 GTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATGGGTGGAA 551
         :::|||::::::         ::::::|||   |||      ||||||
     101 IleThrSerThrThrAsnGlyAsnLeuThrIleTyrSerGlyGlyTrp.. 116

     552 ATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATCGGAGCAGGC. 600
           :::|||::::::|||         ::::::   :::   :::|||
     117 ....ValAspValHisLys.........AsnIleThrLeuLysSerGlyT 129

     601 .................AGACAATATTGGCGGTCTGATGAAGACGAACCC 633
                         :::|||         :::   ||||||:::|||
     129 yrLeuAsnIleThrThrLysGlnGlyAspIleAlaPheGluAspLysPro 145

     634 AAT.........................................AACCG 641
         :::                                        |||::
     146 GlyLeuSerAsnLeuThrIleThrAlaLysGlyThrIleAlaValAsnAs 162

     642 CGAAAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCA 691
         ::::::::::::::::::                          :
     162 nLysLysGlyPheArgPhe............................A 169

     692 ATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGC 741
         :::::   :::   ||||||:::|||||||||
     169 spAsnValThrLeuAsnGlyThrGlyGlyGlyLeu.............. 180

     742 GAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGA........ 783
                           :::   ::::::   ||||||
     181 .................SerPheLysTyrIleGluThrGlyAsnArgAs 191

     784 .......................GGCTCATTTGGCGACAGTGGCTCAC 808
                             |||   ::::::   ||||||:::
     191 pSerAsnPheGluThrHisPheArgGlyArgLeuAsnIleSerGlyLysV 208

     809 CAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTA 858
```


208 alAspIleLeuMetGlnAlaArgGlnGluAsnTrp 219
859 TTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTAGT 908
220 As 220
909 TCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAG 958
220 nArgArgHisTrp GlyArgSerHisTrpA 230
959 TATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAAC 999
230 snValThrArgLeuAsnValSerGluAsnSerTyrPheAsnValThrIle 246
1000 GACAATAATAATGGCGCAGGAAAAATCGATGCCAA 1034
247 AspSerSerGlySerAlaSerSerProGlyAlaGlyProLeuAsnAlaGl 263
1035 ACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAAT 1084
263 nSerGlyLeuAsnGlyIleSerPheAsnAsnAspThr V 276
1085 TGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCT 1134
276 alPheAsnIleAla Ala 281
1135 GCAGGTGGGGTCAAC AGTTATCGACCC AGACTGAA 1169
282 SerSerAlaValAsnPheAsnIleLysProProIleValAspLysValTh 298
1170 TAATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATAC 1219
298 rAsnGlyAsnHisThrLeuPhe LysGlyAsnIleSerV 311
1220 TTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAAT 1269
311 alLeu GlyGlyGlyAspValAsnPhe His 320
1270 TTTACGGTCTCGCCTAAAAACAACGAAACGTGG CAAGG 1307
321 PheAsnAlaSerSerSerAsnTyrGlnThrTyrGlyValIleIleGluSe 337
1308 CGCGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACG 1357
337 rGlnAsnPheSerAlaSerGlyGlySerSerLeuLysPheLysSerGluG 354
1358 GCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTT 1407
354 lySerThrHisAlaAlaPheThrIle LysAsnAspLeuIleLeu 368
1408 CAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGT 1457
369 AsnAlaThrGly GlyAsnIl 375
1458 CATCTTAGATCAGCAGGCG GACGATCAAGGCAAAAAACAAGCCT 1501
375 eSerLeuAsnGlnValAlaGlyIleAspSerAsnLeuLysLysSerLeuI 392
1502 TTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCC 1551
392 leAlaAsnLysAsnIleThrPheGluGlyGlyAsnIleThrLeuAlaAla 408
1552 GATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACG 1601
409 AspLysLysProIleGluIleLysGlyAsnIl 419
1602 TTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAA AATA 1648
419 eThrValLysGluGlyAlaAsnValThrLeuArgSerAlaAsnTyrGlyA 436
1649 CCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACC 1698


436 snAspLysSerAlaLeuSerIleArgGlyAsnValThrAsnLysGlyAsn 452
1699 GTTACCATTACAGGCAAT AAAGATATTACTACAAC 1733
453 LeuThrValThrGlySerAlaIleAsnIleGluLysAsnLeuThrValGl 469
1734 CGGCAAT 1740
469 uGlySerAlaLysPheLeuAlaAsnProAsnTyrSerPheAsnValSerG 486
1741 AACAACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGT 1782
486 lyLeuPheAspAsnGlnGlyLysSerAsnIleSerIleAlaLysGlyGly 502
1783 TGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAA 1832
503 AlaIle PheLysAspIleGluAsnThr GlySerLeuAsnIleTh 517
1833 TTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAA 1882
517 rThrLysSerAspSerAsnHisHisThrIleIleLysGlyAsnIleThrA 534
1883 ATTTAAACGGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGC 1932
534 snArgLysGlyAspLeuAsnIleThrAsn AsnGly 545
1933 AGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAAT 1982
546 AspAsnThrGlu IleGlnIleGlyGlyAsnIleSerGlnLy 559
1983 GGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACC 2032
559 sGlu GlyAsnLeuThrIleSerSerAspLysValAsnI 572
2033 GCACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGA 2070
572 leThr GluArgIleThrIleLysAlaGlyValAsnGlyAsp 585
2071 CAAGCGGTGGTTTCCCGCAATGTTGCC 2097
586 AsnSerAspSerAsnGluAlaThrSerAlaAsnLeuThrIleLysThrLy 602
2098 AAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCAG 2140
602 sGluLeuLysLeuThrAsnAspLeuAsnIleSerGlyPheAsnLysAlaG 619
2141 TTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGAC 2190
619 luIleThrAlaLysAspAsnSerAsnLeuThrIleGlyAspAsnSerAsp 635
2191 TGG ACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGACGA 2234
636 AlaGlyAsnThrAspAlaLysLysValThrPheSerAsnValLysAspSe 652
2235 TAAAGTGATTGCTTCA TTGAGCAAGACCGACA 2266
652 rLysIleSerAlaSerAspHisAsnValThrLeuAsnSerLysValGluT 669
2267 TCAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGA 2316
669 hrSerGlyAspThrAspSerThrGluAspGlyGlyAsnAsnAsnThrGly 685
2317 CTT GCCACACTCAACGGCAATCTTAGTGCAGG 2348
686 LeuThrIleThrAlaLysAsnValThrValAsnAsnAsnIleThrSerHi 702
2349 CGGAGACACGCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACC 2398
702 sLysThrValAsnIleThrAlaSerGluAsnValThrThrLysAlaGlyT 719
2399 TCAGCCTC GTGGGCAATGCCCAAGCAACATTTAATCAAGCC 2439
719 hrThrIleAsnAlaThrThrGlySerValGluValThrAlaLysThrGly 735

```
2440 ACATTAAACGGC.........................AACACATCGGC 2462
        ::::::|||                          |||   :::||
 736 AspIleLysGlyGlyIleGluSerAsnSerGlyAsnValAsnIleThrAl 752

2463 TTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGCA 2512
     ||||   :::   ::::::|||:::||||||   :::   ||||||   :
 752 aSerGlyAsp...ThrLeuAsnValSerAsnIleThrGlyGlnAsnValT 768

2513 GTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTC 2562
     ::::::::   ::::::::::|||   ::::::::::::   |||:::::
 768 hrValAlaAlaAlaSerGlyAlaValThrThrThrLysGlySerThrIle 784

2563 AAC.........GGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTT 2603
     |||         ||||||:::::::::::::   |||
 785 AsnAlaThrThrGlyAsnAlaAsnIleThrThrLys.............. 796

2604 TGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCA. 2652
                  ||||||:::|||:::|||          :::|||
 797 ..............ThrGlyGluIleAsnGlyGluValLysSerAlaS 808

2653 ........TTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACG 2694
             :::::::::      |||      ||||||
 808 erGlyAsnValAsnIleThrAlaSerGlyAsnThrLeu............ 820

2695 GAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGC 2744
     :::::::::|||::::::   :::|||:::|||:::|||   ||||||::
 821 AsnValSerAsnIleThrGlyGlnAsnValThrValThrAlaAsnSerGl 837

2745 CTATCGACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGC 2794
     :            |||      :::      |||||||:::   :::|||
 837 y............AlaIleThrThrThrGluGlySerThrIleAsnAla. 849

2795 CGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCG 2844
                                                |||:::
 850 ..........................................ThrThr 851

2845 GCAGAATCCCGTTTCAACACGCTGACG......GTAAACGGCAAATTGAA 2888
     ::::::::::::::::::|||   |||      :::||||||||||:::::
 852 GlyAspAlaAsnIleThrThrGlnThrGlyAsnIleAsnGlyLysValGl 868

2889 CGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCG 2938
     :::::::                                      ||||
 868 uSerSer.....................................SerG 872

2939 GCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTC 2988
     ||::::::   |||      :::::::::          ||||||||||||
 872 lySerValThrLeuIleAlaThrGlyGln.........ThrLeuAlaVal 885

2989 AACAATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGGA 3038
     :::|||   :::::::::   |||:::         :::|||:::   ::
 886 GlyAsnIleSerGlyAspThrValThr.........IleThrAlaAspLy 899

3039 AGGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAA 3088
     :||||||   :::|||      |||:::::::::|||   |||    :::
 899 sGlyLysLeuThrThrGlnThrSerSerLysIleAsnGlyThrLys.Ser 915

3089 ACGAACACGTCGATGCCGGCG.........CATGGCGTTATCAGCTTATC 3129
     ::::::||||||      :::          |||:::||||||  ::
 916 ValThrThrSerSerGlnSerGlyAspIleSerGlyThrIleSerGlyAs 932

3130 CGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAA....CAAG 3175
     ::::   ::::::|||:::::::      :::   ::::::      ::::
 932 nThrValSerValSerAlaThrGlySerLeuThrThrGlnAlaGlySerL 949

3176 AGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCC...... 3219
     ::::::::   |||   |||:::|||:::   |||:::|||:::
 949 ysIleGluAlaLysThrGlyGluAlaAsnValThrSerAlaThrGlyThr 965

3220 ..............................TTGACGGCAAAACAGGC 3236
                                    :::|||||||:::
 966 IleGlyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnThrAs 982

3237 ACAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTG 3286
       :::|||:::   |||:::   |||:::          :::   :::
 982 pAsnLeuThrIleLysAspGlyAlaArgIleLysAlaThrGlyGlyAlaV 999

3287 ACGCGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGT... 3333
       :::|||   |||:::|||   :::   |||   ::::::::::|||
 999 alThrLeuThrAlaThrGlyGlyThrLeuThrThrGluThrSerSerAsp 1015

3334 ............................GTTGCCGAACCGGCCCGGCA 3353
                                    :::|||:::   :::
1016 IleThrSerSerAsnGlyGlnThrThrLeuThrAlaLysAspSerSerIl 1032

3354 GGCAGGCGGGGAAAATGCC.......................GGCATTA 3379
       ||||||:::   ||||||                     |||   :
1032 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrL 1049

3380 TGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCC 3429
     ::   :::          :::|||   ::::::|||   :::   |||
1049 euThrThrValAlaGlySerLys...IleGluAlaAlaSerGlyThrLeu 1064

3430 TTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCC 3479
     :::   :::   ::::::|||:::         |||::::::::::
1065 ValIleAsnAlaLysAspAlaGlnLeuAspGlyAlaAlaSerGly..... 1079

3480 CCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGC 3529
     .
1079 ..................................................... 1079

3530 AACCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTG 3579
                 ::::::   :::::::::          :::|||
1080 ...........AspHisThrValValAsnAlaThrAsnAlaAsnGlySer 1092

3580 AGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATT 3629
     ::::::      ||||||   :::|||          :::      :::||
1093 GlySerValIleAlaThrThrSerSerArgValAsnIleThrGlyAspLe 1109

3630 GGAC..............CGCGTGTTTGCCGAAGACCGCCGCAACGCCG 3664
     |                  ::::::::::::::::   :::|||:::|
1109 uIleThrIleAsnGlyLeuAsnIleIleSerLysAsnGlyLysAsnThrV 1126

3665 TTTGGACAAGCGGCATCCGG...GACACCAAACACTACCGTTCG...... 3705
     ||      :::|||::::::   |||:::|||:::   :::
1126 alLeuLeuLysGlyValGluIleAspValLysTyrIleGlnProGlyIle 1142

3706 .....................CAAGATTTCCGCGCCTACCGCCAACAAAC 3734
                          :::      ||||||   ::::::
1143 AlaSerValAsnGluValIleGluAlaLysArgAlaLeuGluLysValLy 1159

3735 CGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCG 3784
       |||||| :::   :::      :::      :::      |||:::|||
1159 sAspLeu.SerAspGluGluArgGluThrLeuAlaLysLeuGlyValSer 1175

3785 GCATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACGGCATC 3834
     |||      |||                 |||      ::::::   :::
1176 AlaValArgPhe............AlaGluProAsnAsnAlaIleThrIl 1188

3835 GGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCAT 3884
        |||:::::::   :::   |||   |||:::|||      |||   |
1188 eAsnThrGlnAsnGluPheThrThrArgProLeuSerGlnValThrIleS 1205

3885 CGGCAGGT 3892
     ||   |||
1205 erGluGly 1207
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41731

seq_documentation_block:

```
ID   AAR41731 standard; Protein; 1338 AA.
XX
AC   AAR41731;
XX
DT   26-APR-1994  (first entry)
XX
```

```
DE   High molecular weight protein 3 (HMW3).
XX
KW   HMW; high molecular weight protein; virus; vaccine; influenza;
KW   epitope; immunity; haemophilus influenzae.
XX
OS   Haemophilus influenzae.
XX
PN   WO9319090-A.
XX
PD   30-SEP-1993.
XX
PF   16-MAR-1993;   93WO-US02166.
XX
PR   16-MAR-1992;   92GB-0005704.
XX
PA   (BARE/) BARENKAMP S J.
PA   (INRM ) INSERM INST NAT SANTE & RECH  MEDICALE.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1993-320683/40.
DR   N-PSDB; AAQ49510.
XX
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties
XX
PS   Claim 5; Figure 10; 100pp; English.
XX
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.
XX
SQ   Sequence   1338 AA;


alignment_scores:
           Quality:  308.50            Length:   1542
             Ratio:  0.439               Gaps:     75
 Percent Similarity:  45.525   Percent Identity: 20.169

alignment_block:
 US-09-303-518D-653 x AAR41731   ..

 Align seg 1/1  to: AAR41731  from: 1  to: 1338

       1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCTAA 50
         |||   |||   ||||||   |||   :::|||
      67 MetIleThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLe 83

      51 AACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGT 100
          :::|||:::
      83 uSerGlyLys..................................... 86

     101 TCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATCAAC 150
                           :::::::|||:::|||||||:::|||
      87 ....................GluGlyGlyGluThrTyrLeuGly...... 94

     151 TACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGG 200
                             ::::::::::::|||:::         ||
      95 ....................GlyAspGluArgGlyGlu.........Gl 101

     201 GGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAAT 250
         |   :::   |||::::::   :::|||                   :
     101 yLysAsnGlyIleGlnLeuAlaLysLys.................ThrT 112

     251 CGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAACGGC 300
         :::::   |||:::      |||:::   |||      ::::::::::|||
     112 hrLeuGluLysGlySerThrIleAsnValSerGlyLysGluLysGlyGly 128

     301 GTGGCGGCATTGGCGGGCGATCAATATATTGTG................. 333
           |||   :::   ||||||      ::::::

     129 ArgAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAl 145

     334 .........AGCGTGGCACATAACGGCGGC.................... 354
                 ::::::|||   :::|||||||
     145 aGlnGlyLysAspIleAlaLysThrGlyGlyPheValGluThrSerGlyH 162

     355 ..TATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCC......... 393
           |||   :::::::|||   :::|||      ::::::
     162 isTyrLeuSerIleAspAspAsnAlaIleValLysThrLysGluTrpLeu 178

     394 ...GATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAA 440
            |||   ::::::   :::   :::      :::   ::::::   ::
     179 LeuAspProGluAsnValThrIleGluAlaProSerAlaSerArgValGl 195

     441 AGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTT 490
         :   |||:::::::   :::   ::::::::::::                |
     195 uLeuGlyAlaAspArgAsnSerHisSerAlaGluValIleLysValThrL 212

     491 TGCACAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATG 540
         ||   |||      |||:::         :::      |||:::
     212 euLysLysAsnAsnThrSerLeuThrThrLeuThrAsnThrThrIleSer 228

     541 GATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAAT 590
         :::      |||   |||   :::   :::         |||   :::::
     229 AsnLeuLeuLysSerAlaHisValValAsnIleThrAlaArgArgLysLe 245

     591 C.................................................. 591
         :
     245 uThrValAsnSerSerIleSerIleGluArgGlySerHisLeuIleLeuH 262

     592 ........GGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAACCC 633
                 |||   |||   |||      :::   |||:::|||
     262 isSerGluGlyGlnGlyGlyGlnGlyValGlnIleAspLysAsp...... 276

     634 AATAACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGT 683
                               |||:::|||
     277 .......................IleThrSerGlu.............. 280

     684 CGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT............. 720
          |||||||||                |||||||||
     281 .GlyGlyAsnLeuThrIleTyr......SerGlyGlyTrpValAspValH 295

     721 ..GGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGT 768
         .    :::::::::|||||||||                        |||
     295 isLysAsnIleThrLeuGlySer.......................Gly 303

     769 TTTTTACCAACAGGAGGCTCATTTGGCGAC.................AG 800
         ||||||           :::   ||||||                 ||
     304 PheLeuAsnIleThrThrLysGluGlyAspIleAlaPheGluAspLysSe 320

     801 TGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTA 850
         ||||         :::         ||||||
     320 rGlyArgAsnAsnLeuThrIleThrAlaGln................... 330

     851 ATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 900
           |||::::::   :::||||||                |||||||||||||
     331 ..GlyThrIleThrSerGlyAsn..............SerAsnGlyPhe 341

     901 CAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATAC 950
         :::                        ::::::::::::      ::::::
     342 Arg.....................PheAsnAsnValSerLeuAsnSerLe 351

     951 CCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAACG 1000
                                        :::||||||      |||:::|
     351 u..........................GlyGlyLysLeuSerPheThrA 359

    1001 ACAATAATAATGGCGCAGGAAAA........................... 1023
         ||:::::::::      |||:::
     359 spSerArgGluAspArgGlyArgArgThrLysGlyAsnIleSerAsnLys 375

    1024 ................................ATCGATGCCAAACA 1037
                                         |||:::   |||
     376 PheAspGlyThrLeuAsnIleSerGlyThrValAspIleSerMetLysAl 392
```

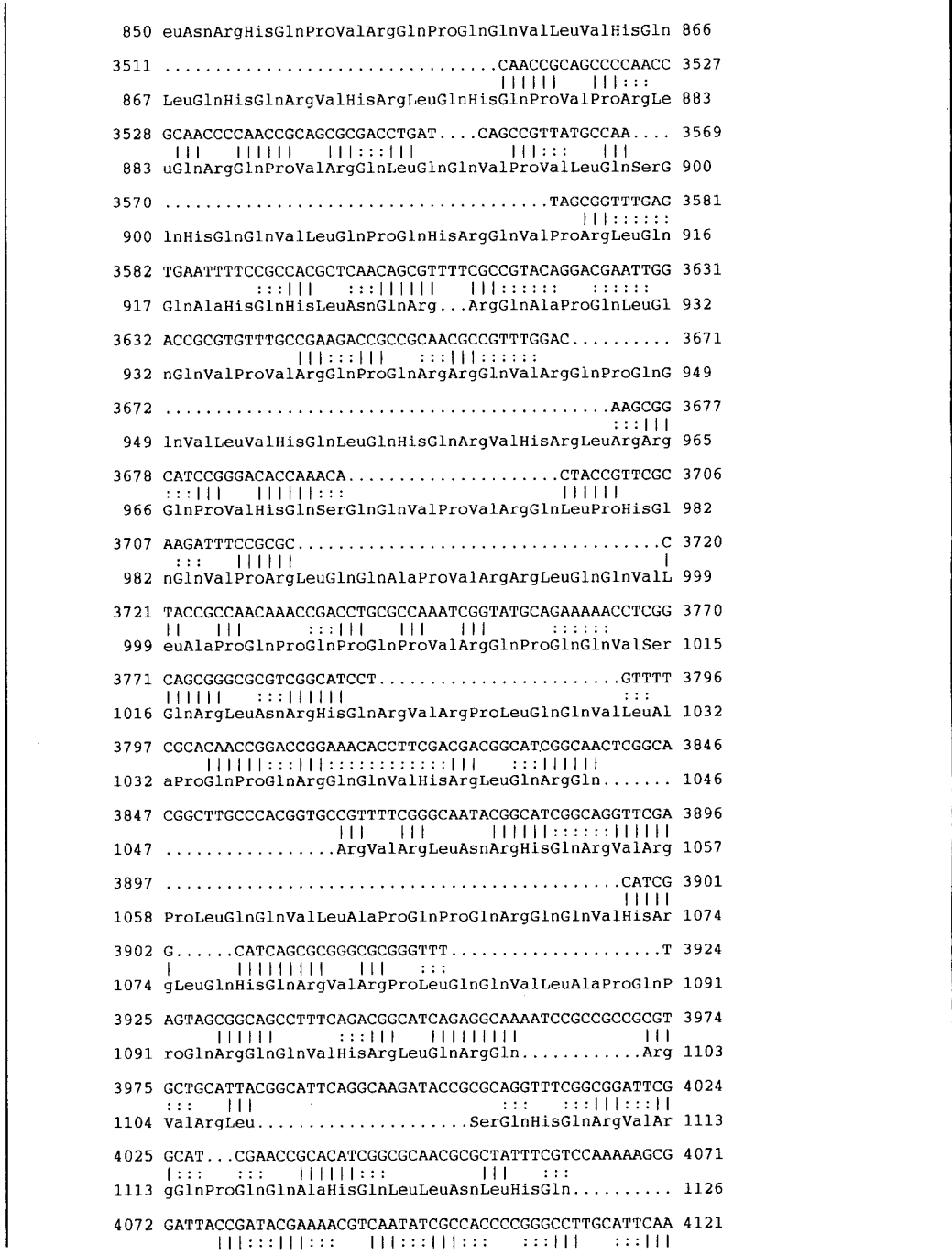
```
     :|||   :::::::|||      :::      |||:::      :::|||
 920 lThrAlaThrGluAsnLeuValThrThrGluAsnAlaValIleAsnAlaT 937

2561 TCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAAC 2610
       :::|||:::|||:::::::::   |||
 937 hrSerGlyThrValAsnIleSerThrLys..................... 946

2611 AGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCA........ 2652
            ||||||   |||:::|||||   :::::::::
 947 .........ThrGlyAspIleLysGlyGlyIleGluSerThrSerGlyAs 960

2653 .TTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAG 2701
      :::::::::      |||      ||||||            :::::::
 960 nValAsnIleThrAlaSerGlyAsnThrLeu............LysValS 973

2702 GCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGA 2751
     ::|||:::::::   :::::::::|||:::|||   :::         .
 973 erAsnIleThrGlyGlnAspValThrValThrAlaAsp............ 985

2752 CACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCG 2801
            |||||||||   |||   :::|||:::
 986 .........AlaGlyAlaLeuThrThrThrAlaGly.............. 994

2802 CCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAAT 2851
               |||   :::|||:::|||      ::::::
 995 .............SerThrIleSerAlaThrThrGlyAsnAla....... 1004

2852 CCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA 2901
         |||      |||   :::|||   :::|||||||:::
1005 ........AsnIleThrThrLysThrGlyAspIleAsnGlyLys...... 1016

2902 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCT 2951

1016 ..................................................... 1016

2952 GGCGGAAAGTTCCGAAGGCACTTAC.....................ACCT 2980
      :::||||||||||:::|||:::                      ||||
1017 .ValGluSerSerSerGlySerValThrLeuValAlaThrGlyAlaThrL 1033

2981 TGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACG 3030
     |||||||||:::|||   :::::::::   |||:::         :::|||
1033 euAlaValGlyAsnIleSerGlyAsnThrValThr.........IleThr 1046

3031 GTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCAC 3080
     :::   :::||||||   ::::::   ::::::::::::::::|||   ||
1047 AlaAspSerGlyLysLeuThrSerThrValGlySerThrIleAsnGlyTh 1063

3081 CCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCC 3130
     |
1063 r..................................................... 1063

3131 GCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTT 3180
                           ::::::   :::   :::::::::
1064 .....................AsnSerValThrThrSerSerGlnSer 1072

3181 TCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAA 3230
     :::|||      |||      :::   :::   :::   :::|||||||:
1073 GlyAspIleGluGlyThrIleSerGlyAsnThrValAsnValThrAlaSe 1089

3231 ACAGGCACAACTT............GCCGCCAAACAACAGGCGGAAAAAG 3268
     :   ::::::|||           :::||||||   :::|||   |||:
1089 rThrGlyAspLeuThrIleGlyAsnSerAlaLysValGluAla...LysA 1105

3269 ACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCAATGCCACC 3318
     :::::|||   :::|||   |||         :::|||:::   :::|||
1105 snGlyAlaAlaThrLeuThrAla......GluSerGlyLysLeuThrThr 1119

3319 GAAAAGGCAGAAAGTGTT............................GC 3338
     :::   :::::::|||:::                            ||
1120 GlnThrGlySerSerIleThrSerSerAsnGlyGlnThrThrLeuThrAl 1136

3339 CGAACCGGCCCGGCAGGCAGGCGGGGAAAATGCC............... 3372
     |:::   :::      ||||||:::   ||||||
```

```
1136 aLysAspSerSerIleAlaGlyAsnIleAsnAlaAlaAsnValThrLeuA 1153

3373 ........GGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCG 3414
             |||   :::   :::      ::::::|||   ::::::|||
1153 snThrThrGlyThrLeuThrThrThrGlyAspSerLys...IleAsnAla 1168

3415 GATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGC 3464
        :::   |||         :::   :::::::|||:::         ||
1169 ThrSerGlyThrLeuThrIleAsnAlaLysAspAlaLysLeuAspGlyAl 1185

3465 TACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAAC 3514
     |:::::::::
1185 aAlaSerGly..................................... 1188

3515 CGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGTTAT 3564
                                    :::|||   :::::::::
1189 .........................AspArgThrValValAsnAlaThr 1196

3565 GCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGC 3614
           ||||||   ::::::   :::|||      :::|||
1197 AsnAlaSerGlySerGlyAsnValThrAlaLysThrSerSerSerValAs 1213

3615 CGTACAGGACGAATTGGAC..............CGCGTGTTTGCCGAAG 3649
      :::      :::|||:::               :::::::::::::|||:
1213 nIleThrGlyAspLeuAsnThrIleAsnGlyLeuAsnIleIleSerGluA 1230

3650 ACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGG...GACACCAAACAC 3696
     ::   ||||||:::|||         |||   :::   |||:::|||:::
1230 snGlyArgAsnThrValArgLeuArgGlyLysGluIleAspValLysTyr 1246

3697 TACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCA 3746
               |||    ::::::  ::::::::::::         ||||
1247 Ile......GlnPro.GlyValAlaSerValGluGluValIleGluAlaL 1261

3747 AATCGGTATGCAGAAAAACC.........................TCG 3769
     ||   |||   :::|||:::                          :::
1261 ysArgValLeuGluLysValLysAspLeuSerAspGluGluArgGluThr 1277

3770 GCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACACC 3819
        |||      ::::::|||:::   |||   :::   |||:::::::
1278 LeuAlaLysLeuGlyValSerAlaValArgPheValGluProAsnAsnAl 1294

3820 TTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTT 3869
          |||:::          |||::::::   :::   |||   |||   |
1294 aIleThrVal......AsnThrGlnAsnGluPheThrThrLysProSerS 1309

3870 CGGGCAATACGGCATCGGCAGGTTCGACA......TCGGCATCAGCGCGG 3913
     ||      |||   |||   |||::::::      |||::::::
1309 erGlnValThrIleSerGluGlyLysAlaCysPheSerSerGlyAsnGly 1325

3914 GCGCGGGTTTTAGTAGCGGCAGCC 3937
     |||||||||   :::   :::|||
1326 AlaArgValCysThrAsnValAla 1333
```

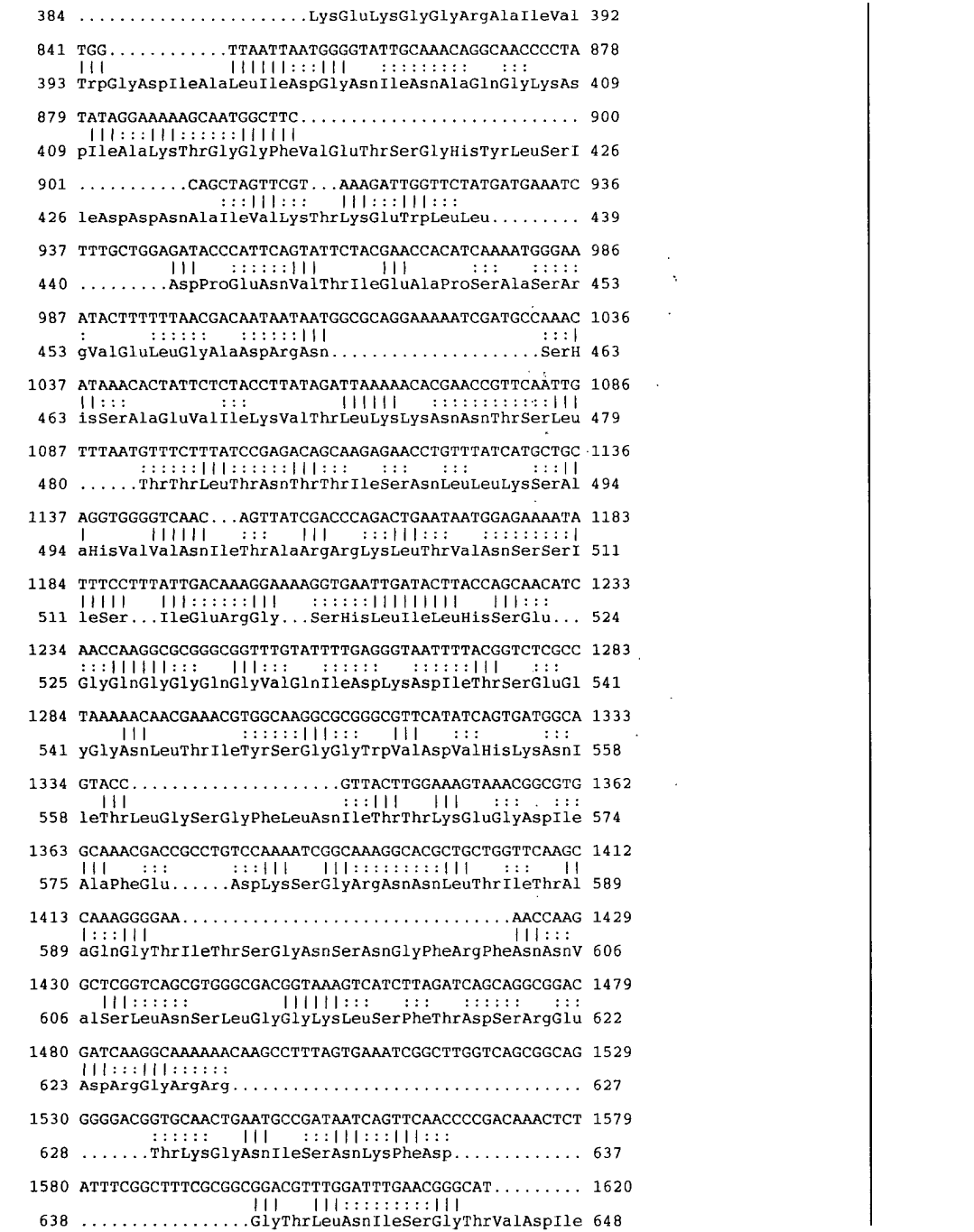
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY81609

seq_documentation_block:
ID AAY81609 standard; Protein; 1237 AA.
XX
AC AAY81609;
XX
DT 24-MAY-2000 (first entry)
XX
DE Streptococcus pneumoniae type 4 protein sequence #109.
XX
KW Streptococcus pneumoniae; vaccine; screening; protein antigen;
KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
KW pneumococcal disease.
XX
OS Streptococcus pneumoniae.
XX
PN WO200006737-A2.
XX

```
PD   10-FEB-2000.
XX
PF   27-JUL-1999;   99WO-GB02451.
XX
PR   27-JUL-1998;   98GB-0016337.
PR   19-MAR-1999;   99US-0125164.
XX
PA   (MICR-) MICROBIAL TECHNICS LTD.
XX
PI   Gilbert CFG,  Hansbro PM;
XX
DR   WPI; 2000-195300/17.
XX
PT   New Streptococcal protein, useful as a vaccine, for diagnosis of
PT   pneumococcal diseases and for screening agents capable of antagonizing
PT   or inhibiting expression of the protein  -
XX
PS   Claim 1; Page 83; 108pp; English.
XX
CC   AAY81501 to AAY81679 represent specifically claimed protein sequences
CC   isolated from Streptococcus pneumoniae. AAA05407 to AAA05590 represent
CC   specifically claimed nucleotide sequences isolated from S. pneumoniae.
CC   The sequences have antibacterial and antiinflammatory properties.
CC   The protein sequences, and fragments of them, are useful as immunogens
CC   and/or antigens. The nucleotide sequences can be used in vaccines and in
CC   diagnostic assays. The proteins and nucleotides can be useful for the
CC   detection and diagnosis of S. pneumoniae. The protein sequences are also
CC   useful for screening an agent capable of antagonising, inhibiting or
CC   interfering with the function or expression of the proteins in which the
CC   agent is useful for treatment or prophylaxis of S. pneumoniae infection
CC   and meningitis. AAA05591 to AAA05614 represent primers used in the
CC   exemplification of the present invention.
XX
SQ   Sequence   1237 AA;


alignment_scores:
           Quality:  307.00            Length:   1099
             Ratio:  0.565               Gaps:     45
 Percent Similarity:  49.409   Percent Identity: 21.110

alignment_block:
 US-09-303-518D-653 x AAY81609   ..

 Align seg 1/1  to: AAY81609  from: 1  to: 1237

    1501 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGC 1550
          ||||||         :::|||:::|||      :::::::::   :::
     186 PheSerGlyAsnThrIleValAsnGlyAlaProAlaIleAsnAlaSerLe 202

    1551 CGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGAC 1600
           :::         :::      |||:::|||            |||
     202 uAsnIleAlaLysSerGluThrLysValTyr.........ThrGlyGluG 216

    1601 GTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACC 1650
           :::|||         :::        :::::::::::::::   |||
     216 lyValAspSerValTyrArgValProIleTyrTyrLysLeuLysValThr 232

    1651 GATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGT 1700
          :::::::|||:::   ::::::                       |||
     233 AsnAspGlySerLysLeuThr.....................PheThrTy 242

    1701 TACCATTACA...GGCAATAAAGATATTACTACAACCGGCAAT..... 1740
           |||:::|||      |||         :::      ||||||
     242 rThrValThrTyrValAsnProLysThrAsnAspLeuGlyAsnIleSerS 259

    1741 .........................AACAACAACTTGGATAGCAAAAAA 1764
                                 |||:::::::   :::::::::
     259 erMetArgProGlyTyrSerIleTyrAsnSerGlyThrSerThrGlnThr 275

    1765 GAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACG.. 1812
           ::::::   ::::::   :::|||:::   :::::::::|||:::
     276 MetLeuThrLeuGlySerAspLeuGlyLysProSerGlyValLysAsnTy 292

    1813 .............AACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAG 1849
                       |||||||||      |||:::|||:::   :::
     292 rIleThrAspLysAsnGlyArgGlnValLeuSerTyrAsnThrSer.... 307

    1850 CGGATCGCACTTTACTGCTTTCCGGCGGAACAAATTTAAACGGCAATATC 1899
                ||||:::      :::|||:::            ||||||
     308 ........ThrMetThrThrGlnGlySerGlyTyrThrTrpGlyAsnGly 321

    1900 ACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGC 1949
          :::|||   ||||||      |||||||:::   :::            ::
     322 AlaGlnMetAsnGly......PhePheAlaLysLys............Gl 332

    1950 CTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCA...C 1996
          :|||:::   |||   |||:::|||:::            :::|||
     332 yTyrGly...LeuThrSerSerTrpThr............ValProIleT 344

    1997 AAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACATTTAAAGCG 2046
           |||           |||::::::::::            :::   |||
     344 hrGlyThr.........AspThrSerPheThrPheThrProTyrAlaAla 357

    2047 GAA..................AACTTCCATATTCAGGGCGGACAAGCGGT 2078
          :::                  |||:::         ||||||::::::||
     358 ArgThrAspArgIleGlyIleAsnTyrPheAsnGlyGlyGlyLysValVa 374

    2079 GGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATC 2128
          |   |||   ::::::::::::                  ||||||:::
     374 lGluSerSerThrThrSerGln..............SerLeuSerGlnS 386

    2129 ACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGT 2178
           :::         :::::::::         |||   :::
     386 erLysSerLeuSerValSerAlaSerGlnSerAlaSerAlaSerAlaSer 402

    2179 ACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTAC 2228
          |||   ::::::   ::::::   ||||||:::::   ::::::   ||
     403 ThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerTh 419

    2229 CGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATG 2278
          |:::   ::::::   ||||||   |||                 ::::::|
     419 rSerAlaSerAlaSerAlaSerThrSer..............AlaSerV 431

    2279 TCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACACTC 2328
          |||||   :::      |||      :::   :::      |||:::
     431 alSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAla 447

    2329 AACGGCAATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAA 2378
          :::::::::   ||||||:::   ::::::      :::::::::   ::
     448 SerAlaSerThrSerAlaSerGluSerAlaSerThrSerAlaSerAlaSe 464

    2379 CGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACAT 2428
          :|||:::   :::::::::   |||         :::|||:::|||:::
     464 rAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerA 481

    2429 TTAAT...CAAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCT 2475
           :::   :::|||:::   :::::::::||||||||||||   :::|||
     481 laSerThrSerAlaSerGluSerAlaSerThrSerAlaSerAlaSerAla 497

    2476 TCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTC 2525
          |||   :::   |||:::::::|||      ::::::|||   :::   ||
     498 SerThrSerAlaSerGluSerAlaSerThrSerAlaSerAlaSerAlaSe 514

    2526 CGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCT 2575
          |   :::|||:::|||::::::|||   ||||||   :::|||:::::::|
     514 rThrSerAlaSerAlaSerAlaSerThrSerAla...SerGlySerAlaS 530

    2576 CCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGA 2625
          ||   :::   :::|||           :::   |||      ::::::
     530 erThrSerThrSerAlaSerAlaSerThrSerAlaSerAla...SerAla 545

    2626 AAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATG 2675
          :::   |||::::::   ::::::::::::      :::::::|||
     546 SerThrSerAlaSerAlaSerAlaSerIleSerAlaSerGluSerAlaSe 562
```



```
    1127 .....ProValArgGlnProGlnHisArgGln...AlaProGlnLeuGln 1140

    4122 CCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAAC 4171
         :::|||
    1141 GlnValPro..................................... 1143

    4172 ACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCC 4221
                                                        |
    1144 .............................................ValA 1145

    4222 GGCAAAGTCCGAACGCGCGTCAATACCGCCGTATT...GGCGCAGGATTT 4268
         |||||   |||:::   ||||||   ||||||:::      :::
    1145 rgGln...ProGlnArgArgGlnValArgArgLeuGlnGlnValProVal 1160

    4269 CGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAA 4313
         ||||||      |||      :::   :::   :::|||:::|||
    1161 ArgGlnProGlnGlnValProValArgGlnProGlnArgArgGln 1175

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30291

seq_documentation_block:
ID   AAW30291 standard; Protein; 1598 AA.
XX
AC   AAW30291;
XX
DT   14-APR-1998  (first entry)
XX
DE   Non-typeable Haemophilus high mol.wt. surface protein HMW3.
XX
KW   Non-typeable Haemophilus; high molecular weight surface protein;
KW   HMW3; immunogen; vaccine; otitis media.
XX
OS   Haemophilus influenzae strain 5.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 113
FT                   /note= "encoded by GTG"
FT   Misc-difference 864
FT                   /note= "encoded by TGT"
XX
PN   WO9736914-A1.
XX
PD   09-OCT-1997.
XX
PF   01-APR-1997;   97WO-US04707.
XX
PR   01-APR-1996;   96US-0617697.
XX
PA   (BARE/) BARENKAMP S J.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1997-503038/46.
DR   N-PSDB; AAT90992.
XX
PT   High molecular weight proteins of non-typeable Haemophilus
PT   influenzae - useful for vaccine production
XX
PS   Claim 1; Page 93-97; 183pp; English.
XX
CC   This protein comprises the high molecular weight surface protein
CC   HMW3 (125 kDa) of non-typeable Haemophilus influenzae strain 5 that
CC   has the immunological ability to protect against disease caused by
CC   a non-typeable Haemophilus strain and is characterised by at least
CC   one surface-exposed B-cell epitope that is recognised by monoclonal
CC   antibody AD6.  The HMW3 amino acid sequence was deduced from an
CC   isolated hmw3 gene (see AAT90992).  HMW1 (see AAW30293), HMW2 (see
CC   AAW30294) and HMW4 (see AAW30292) have also been identified.  A
CC   conjugate comprising HMW3 linked to an antigen, hapten or
CC   polysaccharide, and a synthetic peptide of 6-150 amino acids
CC   corresponding to at least protective epitope of HMW3 are also
CC   claimed.  HMW proteins, conjugates and peptides can be used in
CC   vaccines, as immunogens for preparation of antibodies and as
CC   antigens for detection of these antibodies.
XX
SQ   Sequence   1598 AA;


alignment_scores:
           Quality:  306.00              Length:   1492
             Ratio:  0.435                Gaps:     72
 Percent Similarity:  47.118   Percent Identity: 19.973

alignment_block:
 US-09-303-518D-653 x AAW30291   ..

 Align seg 1/1  to: AAW30291  from: 1  to: 1598

     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
         |||   |||   ::::::||||||   ||||||   :::|||
     199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal..... 211

     228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
              :::|||:::|||   ::::::   :::
     212 .......AsnLeuIleGlyGlyLysValLysAsnGlu..... 221

     278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
           :::|||:::|||   ||||||      ::::::   ||||||||||:::
     222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

     322 CAA.....................................TATATTGTGAG 335
         :::                                     |||   :::::
     238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

     336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
         ::::   :::|||   :::   |||      |||:::   |||:::|||:
     254 aAlaProGluAsnGluAlaIleAsnLeuGlyAspIlePheAlaLysGlyG 271

     386 GCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAAT 435
         ::|||                                       |||
     271 lyAsn.....................................IleAsn 274

     436 TATAAAGCAGGGACT.........AACGGCCATCCTTATGGCGGCGATTA 476
           :::|||:::|||         :::|||         :::   :::
     275 ValArgAlaAlaThrIleArgAsnLysGlyLysLeuSerAlaAspSerVa 291

     477 TCATATGCCGCGTTTGCACAAATTTGTCACAGATGCAGAACCTGTTGAG. 525
                 :::      :::::::|||   :::|||:::      |||
     291 lSerLysAspLysSerGlyAsnIleValLeuSerAlaLysGluGlyGluA 308

     526 .................ATGACCAGTTATATGGATGGGTGGAAATACGCT 558
                           :::::::::      :::      |||   :::
     308 laGluIleGlyGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGly 324

     559 GATTTAAATAAATACCCTGATCGTGTTCGAATCGGAGCAGGCAGACAATA 608
           |||             |||:::|||   :::   :::|||
     325 LysLeuMetIleThrGlyAspLysValThrLeuLysThrGlyAlaValIl 341

     609 TTGGCGGTCTGATGAAGACGAACCCAATAACCGCGAAAGTTCATATCATA 658
               |||   ::::::
     341 eAspLeuSerGlyLysGlu.............................. 347

     659 TTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAATACCTTTGCACAAAAT 708
                                 ||||||:::|||:::
     348 ..........................GlyGlyGluThrTyrLeu...... 353

     709 GGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAG 758
               ||||||          :::   |||::::::   |||:::   ::
     354 ......GlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAl 368

     759 CCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCAC 808
         :             |||      |||:::::::::::::   ||||||
     368 aLysLysThrThrLeuGluLysGlySerThrIleAsnValSerGly.... 383

     809 CAATGTTTATCTATGATGCCCAAAAGCAAAAG.................. 840
                                 |||:::|||                 -
```


384 .....LysGluLysGlyGlyArgAlaIleVal 392
841 TGG.....TTAATTAATGGGGTATTGCAAACAGGCAACCCCTA 878
393 TrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAlaGlnGlyLysAs 409
879 TATAGGAAAAAGCAATGGCTTC..... 900
409 pIleAlaLysThrGlyGlyPheValGluThrSerGlyHisTyrLeuSerI 426
901 ..........CAGCTAGTTCGT...AAAGATTGGTTCTATGATGAAATC 936
426 leAspAspAsnAlaIleValLysThrLysGluTrpLeuLeu..... 439
937 TTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAA 986
440 .....AspProGluAsnValThrIleGluAlaProSerAlaSerAr 453
987 ATACTTTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAAC 1036
453 gValGluLeuGlyAlaAspArgAsn.....SerH 463
1037 ATAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTG 1086
463 isSerAlaGluValIleLysValThrLeuLysLysAsnAsnThrSerLeu 479
1087 TTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGC 1136
480 ......ThrThrLeuThrAsnThrThrIleSerAsnLeuLeuLysSerAl 494
1137 AGGTGGGGTCAAC...AGTTATCGACCCAGACTGAATAATGGAGAAAATA 1183
494 aHisValValAsnIleThrAlaArgArgLysLeuThrValAsnSerSerI 511
1184 TTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATC 1233
511 leSer...IleGluArgGly...SerHisLeuIleLeuHisSerGlu... 524
1234 AACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCC 1283
525 GlyGlnGlyGlyGlnGlyValGlnIleAspLysAspIleThrSerGluGl 541
1284 TAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGGCA 1333
541 yGlyAsnLeuThrIleTyrSerGlyGlyTrpValAspValHisLysAsnI 558
1334 GTACC.....GTTACTTGGAAAGTAAACGGCGTG 1362
558 leThrLeuGlySerGlyPheLeuAsnIleThrThrLysGluGlyAspIle 574
1363 GCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGC 1412
575 AlaPheGlu.....AspLysSerGlyArgAsnAsnLeuThrIleThrAl 589
1413 CAAAGGGGAA.....AACCAAG 1429
589 aGlnGlyThrIleThrSerGlyAsnSerAsnGlyPheArgPheAsnAsnV 606
1430 GCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGAC 1479
606 alSerLeuAsnSerLeuGlyGlyLysLeuSerPheThrAspSerArgGlu 622
1480 GATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAG 1529
623 AspArgGlyArgArg..... 627
1530 GGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCT 1579
628 .....ThrLysGlyAsnIleSerAsnLysPheAsp..... 637
1580 ATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCAT..... 1620
638 .....GlyThrLeuAsnIleSerGlyThrValAspIle 648
1621 TCGCTTTCG.....TTCCACCGCATTCAAAATACCGA 1652
649 SerMetLysAlaProLysValSerTrpPheTyrArg.....As 661
1653 TGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTA 1702
661 pLysGlyArgThrTyrTrpAsn.....ValThrThrLeuA 673
1703 CCATTACAGGC.....AATAAAGATATTACTACAACCGGCAAT 1740
673 snValThrSerGlySerLysPheAsnLeuSerIleAspSerThrGlySer 689
1741 AACAACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGT..... 1782
690 GlySerThrGlyProSerIleArgAsnAlaGluLeuAsnGlyIleThrPh 706
1783 .....TGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGC 1822
706 eAsnLysAlaThrPheAsnIleAlaGlnGlySerThrAlaAsnPheSerI 723
1823 TCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTT... 1869
723 leLysAlaSerIleMetProPheLysSerAsnAlaAsnTyrAlaLeuPhe 739
1870 .....TCCGGCGGAACAAATTTAAACGGCAATATCAC 1901
740 AsnGluAspIleSerValSerGlyGlyGlySerLeuAsnPheLysLeuAs 756
1902 GCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCT 1951
756 nAlaSerSerSerAsnIle.....GlnThrPro..... 765
1952 ACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGA 2001
766 .....Gly 766
2002 GAAATCGTGTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGAAAA 2051
767 ValIleIle.....LysSerGlnAs 773
2052 CTTCCATATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAG 2101
773 nPheAsnValSerGlyGlySerThrLeuAsnLeuLysAlaGluGlySerT 790
2102 TGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTC 2151
790 hrGluThrAlaPheSerIleGluAsnAspLeuAsn..... 801
2152 GCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCT 2201
801 801
2202 GACAAGTTGTACCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCAT 2251
802 .LeuAsnAlaThrGlyGlyAsnIleThr.....IleArgGlnV 814
2252 TGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGATCACGCTCAT 2301
814 alGluGlyThrAspSerArgValAsnLysGlyValAlaAlaLysLysAsn 830
2302 TTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGCAGGCGG 2351
831 IleThrPheLysGly.....GlyAsnIleThrPheGlySe 842
2352 AGACACGCACTATACG...GTTACGCGCAACGCCACCCAAAACGGCAACC 2398
842 rGlnLysAlaThrThrGluIleLysGlyAsnValThrIleAsnLysAsn. 858
2399 TCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAAC 2448
859 .....ThrAsnAlaThrLeuArg 864

```
    1361 lGluAlaLysAsnGlyAlaAlaThrLeuThrAlaGlu.....S 1374

    4062 CCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGCCACCCCGGGCC 4111
         ||   |||      |||:::|||   :::||||||:::      ::::::
    1374 erGlyLysLeuThrThrGlnThrGlySerSerIleThrSerSerAsnGly 1390

    4112 TTGCATTCAACCGCTACCGCGCGG 4135
           ::::::|||   ||||||:::
    1391 ...GlnThrThrLeuThrAlaLys 1397

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41732

seq_documentation_block:
ID   AAR41732 standard; Protein; 1529 AA.
XX
AC   AAR41732;
XX
DT   26-APR-1994  (first entry)
XX
DE   High molecular weight protein 4 (HMW4).
XX
KW   HMW; high molecular weight protein; virus; vaccine; influenza;
KW   epitope; immunity; haemophilus influenzae.
XX
OS   Haemophilus influenzae.
XX
PN   WO9319090-A.
XX
PD   30-SEP-1993.
XX
PF   16-MAR-1993;   93WO-US02166.
XX
PR   16-MAR-1992;   92GB-0005704.
XX
PA   (BARE/) BARENKAMP S J.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1993-320683/40.
DR   N-PSDB; AAQ49511.
XX
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties
XX
PS   Claim 6; Figure 10; 100pp; English.
XX
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.
XX
SQ   Sequence   1529 AA;


alignment_scores:
           Quality:  296.50            Length:   1502
             Ratio:  0.436               Gaps:     69
 Percent Similarity:  45.273   Percent Identity: 20.040

alignment_block:
 US-09-303-518D-653 x AAR41732   ..

 Align seg 1/1  to: AAR41732  from: 1  to: 1529

     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
         |||   |||   ::::::||||||   ||||||   :::|||
     127 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal........ 139

     228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
               :::|||:::|||   ::::::   :::
     140 .......AsnLeuIleGlyGlyLysValLysAsnGlu............. 149

     278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
           :::|||:::|||   ||||||      ::::::   |||||||||:::
     150 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 165

     322 CAA.................................TATATTGTGAG 335
         :::                                 |||   :::::
     166 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 182

     336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
         ::::   :::|||   :::   |||      |||:::   |||:::|||:
     182 aAlaProGluAsnGluAlaIleAsnLeuGlyAspIlePheAlaLysGlyG 199

     386 GCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAAT 435
         ::|||
     199 lyAsn.......................................... 200

     436 TATAAAGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCC 485

     200 .......................................... 200

     486 GCGTTTGCACAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTT 535

     200 .......................................... 200

     536 ATATGGATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT 585
                                   :::|||               |||
     201 .........................IleAsn..............Val 203

     586 CGAATCGGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAACCCAA 635
         |||   ::::::   |||:::      :::   :::   |||:::   ::
     204 ArgAlaAlaThrIleArgAsnLysGlyLysLeuSerAlaAspSerValSe 220

     636 TAACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGG......C 679
         ::::   :::|||:::      :::   ||||||   :::
     220 rLysAspLysSerGlyAsnIleValLeuSerAlaLysGluGlyGluAlaG 237

     680 TCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGT...GGTGGCACA 726
           :::||||||         |||||||||   ::::::   ||||||
     237 luIleGlyGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGlyLys 253

     727 GTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGTTTTTTACC 776
         :::   :::   ::::::|||:::                        |||
     254 LeuMetIleThrGlyAspLysValThr.................LeuLy 264

     777 AACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATG 826
          ||||||:::   :::      ||||||:::
     264 sThrGlyAlaValIleAspLeuSerGlyLys..................... 274

     827 CCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCC 876
                                              :::   |||:::
     275 ................................GluGlyGlyGluThr 279

     877 TATATAGGA.................AAAAGCAATGGCTTCCAGCTAGT 908
         |||:::|||                 :::||||||:::|||||||::
     280 TyrLeuGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAl 296

     909 TCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAG 958
         ::::|||                :::      |||:::|||   :::|
     296 aLysLys............ThrThrLeuGluLysGlySerThrIleAsnV 309

     959 TATTCTACGAACCACATCAAAATGGGAAA..................... 987
         ||                  :::|||||||
     309 al.................SerGlyLysGluLysGlyGlyArgAlaIle 319

     988 ..............TACTTTTTTAACGACAATAATAATGGCGCAGGAAA 1022
                        :::::::::   |||   |||:::   |||::
     320 ValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAlaGlnGlySe 336

    1023 AATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACAC 1072
         :      ||||||      :::               ::::::|||
     336 rAspIleAlaLysThrGlyGlyPhe..............ValGluThrS 348
```

```
 825 snIleValGlyAsnLeuSerIleAlaGluAspSer......ThrPheLys 839

2776 GGCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATC 2825
     |||:::|||:::|||                                |||::
 840 GlyGluAlaSerAspAsn.........................LeuAs 847

2826 CGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAA 2875
     ::::|||                              :::|||:::|||   |
 847 nIleThr.........................GlyThrPheThrAsnA 855

2876 ACGGCAAATTGAAC.........GGTCAGGGAACATTCCGCTTTATGTCG 2916
     |||||      |||           ||||||:::   ::::::::::::
 855 snGlyThrAlaAsnIleAsnIleLysGlnGlyValValLysLeuGlnGly 871

2917 GAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGA 2966
     :::::::::      :::::::|||   |||:::::::::   ::::::::
 872 AspIleIle......AsnLysGlyGlyLeuAsnIleThrThrAsnAlaSe 886

2967 AGGCACTTACACCTTGGCTGTCAACAATACC..................... 2997
     :||||||           :::|||::::::
 886 rGlyThrGlnLysThrIleIleAsnGlyAsnIleThrAsnGluLysGlyA 903

2997 ..................................................... 2997

 903 spLeuAsnIleLysAsnIleLysAlaAspAlaGluIleGlnIleGlyGly 919

2998 .................GGCAACGAACCCGTAAGTCTCGAGCAATTGAC 3029
                     ||||||      :::|||   ::::::::::
 920 AsnIleSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysValAs 936

3030 GGTAGTGGAA............................GGAAAAGACA 3049
     ::::::::::                           |||:::::::
 936 nIleThrAsnGlnIleThrIleLysAlaGlyValGluGlyGlyArgSerA 953

3050 ACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTC 3099
     :::::      :::||||||   |||:::|||:::|||
 953 spSerSerGluAlaGluAsnAlaAsnLeuThrIleGln............ 965

3100 GATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCG 3149

 965 ..................................................... 965

3150 CCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGG 3199
                ::::||||||   ::::|||:::      |||:::   :
 966 .............ThrLysGluLeuLysLeuAlaGlyAspLeuAsnIleS 978

3200 CGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCC 3249
     ::|||   :::::::|||   :::|||||||||:::::::::      :::
 978 erGlyPheAsnLysAlaGluIleThrAlaLysAsnGlySerAspLeuThr 994

3250 AAACAACAGGCGGAAAAAGACAACGCGCAAAGC................... 3282
          :::|||:::      ||||||::::::
 995 IleGlyAsnAlaSerGlyGlyAsnAlaAspAlaLysLysValThrPheAs 1011

3283 .......CTTGACGCGCTGATTGCGGCC...GGGCGCAATGCCACCGAAA 3322
             |||:::   |||:::::::   |||:::|||:::|||   :
1011 pLysValLysAspSerLysIleSerThrAspGlyHisAsnValThrLeuA 1028

3323 AGGCAGAA...AGTGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAAT 3369
     :::::|||   ::::::::::::   :::   :::||||||:::::|||
1028 snSerGluValLysThrSerAsnGlySerSerAsnAlaGlyAsnAspAsn 1044

3370 GCC...GGCATTATGCAGGCGGAGGAAGAG..................... 3396
     :::   |||:::      :::   ::::::
1045 SerThrGlyLeuThrIleSerAlaLysAspValThrValAsnAsnAsnVa 1061

3397 .......AAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAAC 3439
             |||   ::::::   :::      ::::::   :::
1061 lThrSerHisLysThrIleAsnIleSerAlaAlaAlaGlyAsnValThrT 1078

3440 AGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGC 3489
       :::|||:::   |||      |||||||||:: |||         ::
1078 hrLysGluGlyThrThrIleAsnAlaThrThrGly.SerValGluValTh 1094

3490 CGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACC 3539
     :|||      |||      :::::::::      ::::::|||   :::
1094 rAlaGlnAsnGlyThrIleLysGlyAsnIleThrSerGlnAsnValThrV 1111

3540 GCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTT 3589
       :::||||||   :::   |||           :::   :::|||
1111 alThrAlaThrGluAsnLeuValThrThrGluAsnAlaValIleAsnAla 1127

3590 CCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTG 3639
               ::::::   :::|||
1128 ThrSerGlyThrValAsnIleSer..................... 1135

3640 TTTGCCGAAGACCGCCGCAACGCCGTTTGGA...CAAGCGGCATCCGGGA 3686
            ||||||:::            |||   :::::::::||||||:
1136 ....ThrLysThrGlyAspIleLysGlyGlyIleGluSerThrSerGlyA 1151

3687 CACCAAACACTACCGTTCGC............AAGATTTCCGCGCCTACC 3724
     ::   |||   |||:::                |||:::|||      |||
1151 snValAsnIleThrAlaSerGlyAsnThrLeuLysValSerAsnIleThr 1167

3725 GCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGC 3774
     ::::::      |||   :::   ::::::          |||:::::||
1168 GlyGlnAspValThrValThrAlaAspAlaGlyAlaLeuThrThrThrAl 1184

3775 GGGCGCGTCGGCATCCTGTTTTCGCACAACCGGA................ 3808
     ||||::::::   |||:::      ||||||||
1184 aGlySerThrIleSerAla......ThrThrGlyAsnAlaAsnIleThrT 1199

3809 ..............................CCGGAAACACCTTCGACG 3826
                                   ::::::   |||:::
1199 hrLysThrGlyAspIleAsnGlyLysValGluSerSerSerGlySerVal 1215

3827 ACGGCATCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAA 3876
     |||      ||||||         |||   :::      :::|||||||||
1216 ThrLeuValAlaThrGlyAlaThrLeuAlaValGlyAsnIleSerGlyAs 1232

3877 T............ACGGCATCGGCAGGTTCGACATCGGCATCAGCGCGGG 3914
     |            ||||||::::::|||:::   ::::::::::::   :
1232 nThrValThrIleThrAlaAspSerGlyLysLeuThrSerThrValGlyS 1249

3915 CGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGAGGCAAAA... 3961
     ::   :::      :::   :::      |||:::|||::::::
1249 erThrIleAsnGlyThrAsnSerValThrThrSerSerGlnSerGlyAsp 1265

3962 ..............TCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCA 3996
                     |||:::   :::      :::||||||
1266 IleGluGlyThrIleSerGlyAsnThrValAsnValThrAlaSerThrGl 1282

3997 AGAT..............ACCGCGCAGGTTTCGGCGGATTCGGCATCGA 4031
      |||               :::|||:::|||:::|||   :::::::::
1282 yAspLeuThrIleGlyAsnSerAlaLysValGluAlaLysAsnGlyAlaA 1299

4032 ACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGAT 4081
           :::|||:::            |||   |||      |||:::
1299 laThrLeuThrAlaGlu............SerGlyLysLeuThrThrGln 1311

4082 ACGAAAACGTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCTACCGC 4131
     |||   :::|||||||:::      ::::::   ::::::|||   |||||
1312 ThrGlySerSerIleThrSerSerAsnGly...GlnThrThrLeuThrAl 1327

4132 GCGG 4135
     |:::
1327 aLys 1328
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30292

seq_documentation_block:
ID AAW30292 standard; Protein; 1601 AA.
XX
AC AAW30292;
XX
DT 14-APR-1998 (first entry)

```
XX
DE   Non-typeable Haemophilus high mol.wt. surface protein HMW4.
XX
KW   Non-typeable Haemophilus; high molecular weight surface protein;
KW   HMW4; immunogen; vaccine; otitis media.
XX
OS   Haemophilus influenzae strain 5.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 372
FT                   /note= "encoded by TCT"
FT   Misc-difference 400
FT                   /note= "encoded by AAT"
XX
PN   WO9736914-A1.
XX
PD   09-OCT-1997.
XX
PF   01-APR-1997;   97WO-US04707.
XX
PR   01-APR-1996;   96US-0617697.
XX
PA   (BARE/) BARENKAMP S J.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1997-503038/46.
DR   N-PSDB; AAT90993.
XX
PT   High molecular weight proteins of non-typeable Haemophilus
PT   influenzae - useful for vaccine production
XX
PS   Claim 1; Page 97-102; 183pp; English.
XX
CC   This protein comprises the high molecular weight surface protein
CC   HMW4 (123 kDa) of non-typeable Haemophilus influenzae strain 5 that
CC   has the immunological ability to protect against disease caused by
CC   a non-typeable Haemophilus strain and is characterised by at least
CC   one surface-exposed B-cell epitope that is recognised by monoclonal
CC   antibody AD6.  The HMW4 amino acid sequence was deduced from an
CC   isolated hmw4 gene (see AAT90993).  HMW1 (see AAW30293), HMW2 (see
CC   AAW30294) and HMW3 (see AAW30291) have also been identified.  A
CC   conjugate comprising HMW4 linked to an antigen, hapten or
CC   polysaccharide, and a synthetic peptide of 6-150 amino acids
CC   corresponding to at least protective epitope of HMW4 are also
CC   claimed.  HMW proteins, conjugates and peptides can be used in
CC   vaccines, as immunogens for preparation of antibodies and as
CC   antigens for detection of these antibodies.
XX
SQ   Sequence   1601 AA;


alignment_scores:
            Quality:  296.50              Length:   1502
              Ratio:  0.436                Gaps:     69
 Percent Similarity:  45.273   Percent Identity: 20.040

alignment_block:
 US-09-303-518D-653 x AAW30292   ..

 Align seg 1/1  to: AAW30292  from: 1  to: 1601

     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
         |||  |||  :::::|||||  ||||||  :::|||
     199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal........ 211

     228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
             :::|||:::|||  ::::::  :::
     212 .......AsnLeuIleGlyGlyLysValLysAsnGlu............. 221

     278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
       :::|||:::|||  ||||||     ::::::  |||||||||:::
     222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

     322 CAA.................................TATATTGTGAG 335
         :::                                 |||  :::::
     238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

     336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
         ::::  :::|||  :::  |||     |||:::  |||:::|||:
     254 aAlaProGluAsnGluAlaIleAsnLeuGlyAspIlePheAlaLysGlyG 271

     386 GCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAAT 435
         ::|||
     271 lyAsn.......................................... 272

     436 TATAAAGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCC 485

     272 .......................................... 272

     486 GCGTTTGCACAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTT 535

     272 .......................................... 272

     536 ATATGGATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT 585
                                  :::|||              |||
     273 ........................IleAsn.............Val 275

     586 CGAATCGGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAACCCAA 635
         |||  ::::::  |||:::     :::  :::  |||:::  ::
     276 ArgAlaAlaThrIleArgAsnLysGlyLysLeuSerAlaAspSerValSe 292

     636 TAACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGG......C 679
         ::::  :::|||:::     :::  ||||||  :::
     292 rLysAspLysSerGlyAsnIleValLeuSerAlaLysGluGlyGluAlaG 309

     680 TCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGT...GGTGGCACA 726
           :::||||||        |||||||||  ::::::  ||||||
     309 luIleGlyGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGlyLys 325

     727 GTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGTTTTTTACC 776
         :::  :::  ::::::|||:::                    |||
     326 LeuMetIleThrGlyAspLysValThr.................LeuLy 336

     777 AACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATG 826
          ||||||:::  :::     ||||||:::
     336 sThrGlyAlaValIleAspLeuSerGlyLys..................... 346

     827 CCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCC 876
                                         :::  |||:::
     347 ................................GluGlyGlyGluThr 351

     877 TATATAGGA.................AAAAGCAATGGCTTCCAGCTAGT 908
         |||:::|||                :::||||||:::||||||::
     352 TyrLeuGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAl 368

     909 TCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAG 958
         ::::|||                :::     |||:::|||  :::|
     368 aLysLys............ThrThrLeuGluLysGlySerThrIleAsnV 381

     959 TATTCTACGAACCACATCAAAATGGGAAA..................... 987
         ||                 :::||||||
     381 al.................SerGlyLysGluLysGlyGlyArgAlaIle 391

     988 ..............TACTTTTTTAACGACAATAATAATGGCGCAGGAAA 1022
                        :::::::::  |||  |||:::  |||::
     392 ValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAlaGlnGlySe 408

    1023 AATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACAC 1072
         :     ||||||        :::              ::::::|||
     408 rAspIleAlaLysThrGlyGlyPhe..............ValGluThrS 420

    1073 GAACCGTTCAATTG..............TTTAATGTTTCTTTATCCGAG 1107
               :::|||                  |||:::  :::|||
     420 erGlyHisAspLeuSerIleGlyAspAspValIleValAspAlaLysGlu 436
```

```
 912 GlyGluAlaSerAspAsn.........................LeuAs 919

2826 CGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAA 2875
     ::::|||                          :::|||:::|||   |
 919 nIleThr.........................GlyThrPheThrAsnA 927

2876 ACGGCAAATTGAAC.........GGTCAGGGAACATTCCGCTTTATGTCG 2916
     |||||      |||           ||||||:::   ::::::::::::
 927 snGlyThrAlaAsnIleAsnIleLysGlnGlyValValLysLeuGlnGly 943

2917 GAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGA 2966
     ::::::::::      :::::::|||   |||::::::::::   ::::::::
 944 AspIleIle......AsnLysGlyGlyLeuAsnIleThrThrAsnAlaSe 958

2967 AGGCACTTACACCTTGGCTGTCAACAATACC..................... 2997
     :||||||           :::|||::::::
 958 rGlyThrGlnLysThrIleIleAsnGlyAsnIleThrAsnGluLysGlyA 975

2997 ................................................... 2997

 975 spLeuAsnIleLysAsnIleLysAlaAspAlaGluIleGlnIleGlyGly 991

2998 .................GGCAACGAACCCGTAAGTCTCGAGCAATTGAC 3029
                     ||||||      ::::||   ::::::::::::
 992 AsnIleSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysValAs 1008

3030 GGTAGTGGAA..........................GGAAAAGACA 3049
     ::::::::::                          |||:::::::
1008 nIleThrAsnGlnIleThrIleLysAlaGlyValGluGlyGlyArgSerA 1025

3050 ACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTC 3099
     :::::      :::|||||   |||:::|||:::|||
1025 spSerSerGluAlaGluAsnAlaAsnLeuThrIleGln............ 1037

3100 GATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCG 3149

1037 ................................................... 1037

3150 CCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGG 3199
                ::::||||||   :::|||:::      |||:::   :
1038 .............ThrLysGluLeuLysLeuAlaGlyAspLeuAsnIleS 1050

3200 CGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCC 3249
     ::|||   :::::::|||   :::|||||||||||:::::::::      :::
1050 erGlyPheAsnLysAlaGluIleThrAlaLysAsnGlySerAspLeuThr 1066

3250 AAACAACAGGCGGAAAAAGACAACGCGCAAAGC................... 3282
           :::|||:::      ||||||:::::
1067 IleGlyAsnAlaSerGlyGlyAsnAlaAspAlaLysLysValThrPheAs 1083

3283 .......CTTGACGCGCTGATTGCGGCC...GGGCGCAATGCCACCGAAA 3322
             |||:::   |||:::::   |||:::|||:::|||   :
1083 pLysValLysAspSerLysIleSerThrAspGlyHisAsnValThrLeuA 1100

3323 AGGCAGAA...AGTGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAAT 3369
     :::::|||   ::::::::::::   :::   :::||||||:::::|||
1100 snSerGluValLysThrSerAsnGlySerSerAsnAlaGlyAsnAspAsn 1116

3370 GCC...GGCATTATGCAGGCGGAGGAAGAG..................... 3396
     :::   |||:::      :::   ::::::
1117 SerThrGlyLeuThrIleSerAlaLysAspValThrValAsnAsnIleVa 1133

3397 .......AAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAAC 3439
             |||   ::::::   :::      ::::::   :::
1133 lThrSerHisLysThrIleAsnIleSerAlaAlaAlaGlyAsnValThrT 1150

3440 AGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGC 3489
       :::|||:::   |||      |||||||||:::  |||         ::
1150 hrLysGluGlyThrThrIleAsnAlaThrThrGly.SerValGluValTh 1166

3490 CGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACC 3539
     :|||      |||      :::::::::      :::::::|||   :::
1166 rAlaGlnAsnGlyThrIleLysGlyAsnIleThrSerGlnAsnValThrV 1183

3540 GCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTT 3589
       :::||||||   :::   |||           :::   :::|||
1183 alThrAlaThrGluAsnLeuValThrThrGluAsnAlaValIleAsnAla 1199

3590 CCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTG 3639
           ::::::   :::|||
1200 ThrSerGlyThrValAsnIleSer........................... 1207

3640 TTTGCCGAAGACCGCCGCAACGCCGTTTGGA...CAAGCGGCATCCGGGA 3686
           ||||||:::            |||   ::::::::::||||||:
1208 ....ThrLysThrGlyAspIleLysGlyGlyIleGluSerThrSerGlyA 1223

3687 CACCAAACACTACCGTTCGC............AAGATTTCCGCGCCTACC 3724
     ::   |||   |||:::               |||:::|||      |||
1223 snValAsnIleThrAlaSerGlyAsnThrLeuLysValSerAsnIleThr 1239

3725 GCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGC 3774
     ::::::      |||   :::   ::::::          |||:::::::||
1240 GlyGlnAspValThrValThrAlaAspAlaGlyAlaLeuThrThrThrAl 1256

3775 GGGCGCGTCGGCATCCTGTTTTCGCACAACCGGA................ 3808
     ||||::::::   |||:::      |||||||||
1256 aGlySerThrIleSerAla......ThrThrGlyAsnAlaAsnIleThrT 1271

3809 ..............................CCGGAAACACCTTCGACG 3826
                                    ::::::   |||:::
1271 hrLysThrGlyAspIleAsnGlyLysValGluSerSerSerGlySerVal 1287

3827 ACGGCATCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAA 3876
     |||      ||||||         |||   :::      :::||||||||
1288 ThrLeuValAlaThrGlyAlaThrLeuAlaValGlyAsnIleSerGlyAs 1304

3877 T............ACGGCATCGGCAGGTTCGACATCGGCATCAGCGCGGG 3914
     |            ||||||::::::|||:::   ::::::::::::   :
1304 nThrValThrIleThrAlaAspSerGlyLysLeuThrSerThrValGlyS 1321

3915 CGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGAGGCAAAA... 3961
     ::   :::      :::   :::      |||:::|||::::::
1321 erThrIleAsnGlyThrAsnSerValThrThrSerSerGlnSerGlyAsp 1337

3962 ..............TCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCA 3996
                     |||:::   :::      :::|||||||
1338 IleGluGlyThrIleSerGlyAsnThrValAsnValThrAlaSerThrGl 1354

3997 AGAT..............ACCGCGCAGGTTTCGGCGGATTCGGCATCGA 4031
      |||               :::|||:::|||:::|||   :::::::::
1354 yAspLeuThrIleGlyAsnSerAlaLysValGluAlaLysAsnGlyAlaA 1371

4032 ACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGAT 4081
           :::|||:::            |||   |||      |||:::
1371 laThrLeuThrAlaGlu............SerGlyLysLeuThrThrGln 1383

4082 ACGAAAACGTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCTACCGC 4131
     |||   :::|||||||:::      ::::::   :::::|||   |||||
1384 ThrGlySerSerIleThrSerSerAsnGly...GlnThrThrLeuThrAl 1399

4132 GCGG 4135
     |:::
1399 aLys 1400
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41728

seq_documentation_block:
ID AAR41728 standard; Protein; 1477 AA.
XX
AC AAR41728;
XX
DT 26-APR-1994 (first entry)
XX
DE High molecular weight protein 2 (HMW2).
XX
KW HMW; high molecular weight protein; virus; vaccine; influenza;

```
KW   epitope; immunity; haemophilus influenzae; gene cluster.
XX
OS   Haemophilus influenzae.
XX
PN   WO9319090-A.
XX
PD   30-SEP-1993.
XX
PF   16-MAR-1993;   93WO-US02166.
XX
PR   16-MAR-1992;   92GB-0005704.
XX
PA   (BARE/) BARENKAMP S J.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1993-320683/40.
DR   N-PSDB; AAQ49509.
XX
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties
XX
PS   Claim 4; Figure 4; 100pp; English.
XX
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.  This sequence is claimed to be
CC   the same as that given in AAR41724 (High molecular weight protein 2)
CC   although it does differ slightly.
XX
SQ   Sequence   1477 AA;
```

```
alignment_scores:
            Quality:  287.00              Length:   1452
              Ratio:  0.407                 Gaps:     69
 Percent Similarity:  48.623   Percent Identity: 20.110

alignment_block:
 US-09-303-518D-653 x AAR41728   ..

 Align seg 1/1  to: AAR41728  from: 1  to: 1477

     374 GTGCGGAGGGAAGCAATC...CCGATCAGCACCGCTTTTCTT..... 412
         ||||||:::   |||:::   |||::|||:::   :::|||
      42 ValArgHisLeuAlaLeuLysProLeuSerAlaMetLeuLeuSerLeuGl 58

     413 ....ACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCA 458
             |||::::::   ::::::::::::   :::   :::|||||||:::
      58 yValThrSerIleProGlnSerValLeuAlaSerGlyAsnLeuThrSerT 75

     459 TCCTTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAG 508
           |||           :::   :::       :::|||   :::
      75 hrLysMetLysTrpCysSerPheTyrLysLysThrSerAsnLysThrIle 91

     509 ATGCAGAACCTGTTGAGATGACCAGTTATATGGATGGGTGGAAATACGCT 558
         ::::::|||   :::         :::   |||         |||
      92 IleArgAsnSerValAspAlaIleIleAsnTrpLysGlnPheAsn..... 106

     559 GATTTAAATAAATACCCTGATCGTGTTC..................... 586
          |||      |||   ::::::   |||
     107 .IleAspGlnAsnGluMetValGlnPheLeuGlnGluAsnAsnAsnSerA 123

     587 .................GAATCGGAGCAGGCAGACAATATTGGCGGTCTG 619
                           |||:::|||   :::::::::   |||:::
     123 laValPheAsnArgValThrSerAsnGlnIleSerGlnLeuLysGlyIle 139

     620 ATGAAGACGAACCCAATAACCGCGAAAGTTCATATCATATTGCAAGCGCA 669
         :::   :::|||         :::::::|||   :::|||
     140 LeuAspSerAsn.........GlyGlnValPheLeuIleAsnProAsnGl 153

     670 TATTCTTGGCTCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGTGG 719
          |||      ::::::   |||||| :::   ||| ||||||      ::
     153 yIleThrIleGlyLysAspAlaIle.IleAsnThr.AsnGlyPheThrAl 169

     720 TGGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGTT 769
         ::::|||:::::::::::::::|||:::||||||
     169 aSerThrLeuAspIleSerAsnGluAsnIleLys..............  180

     770 TTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATC 819
                                                 :::      |||
     181 ..............................AlaArgAsnPheThr 185

     820 TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGG 869
         ::::::      |||:::|||   |||   :::   ::::::::   ||
     186 PheGluGlnThrLysAspLysAlaLeuAlaGlu...IleValAsnHisGl 201

     870 CAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATT 919
         |         :::||||||:::::::::   :::|||:::
     201 yLeuIleThrValGlyLysAspGlySerValAsnLeuIle..........  214

     920 GGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAA 969

     214 ..............................................  214

     970 CCACATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGG 1019
                ::::||||||                          ::::::   ||
     215 .........GlyGlyLysVal..................LysAsnGluGl 222

    1020 AAAAATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAA 1069
         |   |||:::::::::                |||||              |||
     222 yValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGlnLysI 239

    1070 CACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAA 1119
                   ::::::::::|||   :::::::::   :::
     239 leThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAlaAla 255

    1120 CCT.........................GTTTATCATGCTGCAGGTGG 1142
         |||                                ::::::      :::||||::
     256 ProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyGlyAs 272

    1143 GGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCC.... 1188
         ::::|||                  :::::::|||      ::::::|||
     272 nIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSerAlaA 289

    1189 ..TTTATTGACAAAGGAAAA...GGTGAATTGATACTTACCAGCAACATC 1233
             :::::::|||   |||   |||::::::::::|||::::::::::
     289 spSerValSerLysAspLysSerGlyAsnIleValLeuSerAlaLysGlu 305

    1234 AACCAAGGC...GCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTC 1280
         :::::::::      |||||||:::                      :::||
     306 GlyGluAlaGluIleGlyGlyVal.....................IleSe 315

    1281 GCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGATG 1330
         |   :::|||::::::::   :::|||:::   :::   |||:::   |
     315 rAlaGlnAsnGlnGlnAla...LysGlyGlyLysLeuMetIleThr...G 330

    1331 GCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 1380
         ||:::   ||||||   |||   :::|||::::::   |||      :::
     330 lyAspLysValThrLeuLys...ThrGlyAlaValIleAspLeuSerGly 345

    1381 AAAATCGGCAAAGGCACGCTGCTG...GTTCAAGCCAAAGGGGAAAACCA 1427
         |||   |||      |||   |||      :::   :::||||||:::::
     346 LysGluGlyGlyGluThrTyrLeuGlyGlyAspGluArgGlyGluGlyLy 362

    1428 AGGCTCGGTC.................AGCGTGGGCGACGGTAAAGTCA 1459
         ::::::::::                 |||:::      |||:::::::|
     362 sAsnGlyIleGlnLeuAlaLysLysThrSerLeuGluLysGlySerThrI 379

    1460 TCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAA 1509
         ||      :::      :::      |||:::      ::::::::::::
     379 leAsnValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAsp 395
```

```
 903 IleSerGluSerAlaThrPheLysGlyLysThrArgAspThrLeuAsnIl 919

2965 .GAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAA 3013
        |||:::::::|||            |||:::|||:::      ::::
 919 eThrGlyAsnPheThr............AsnAsnGlyThrAlaGluIleA 932

3014 GTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCC 3063
     :::::   |||   :::|||      |||      |||:::      :::
 932 snIleThrGlnGlyValValLysLeuGly......AsnValThrAsnAsp 946

3064 GAAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATG 3113
        :::||||||:::|||   :::   ::::::   :::   :::
 947 GlyAspLeuAsnIleThrThrHisAlaLysArgAsnGlnArgSerIleIl 963

3114 GCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGG 3163
           ::::::|||:::|||   |||:::::::::::::   :::
 963 eGlyGlyAspIleIleAsnLysLysGlySerLeuAsnIleThrAspSerA 980

3164 TCAAAGAACAAGAGCTT.............................. 3180
       ::::::   |||:::
 980 snAsnAspAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysGluGly 996

3181 ..............TCCGACAAACTC..................... 3192
                   |||||||||:::
 997 AsnLeuThrIleSerSerAspLysIleAsnIleThrLysGlnIleThrIl 1013

3193 .......GGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGG 3235
           |||      ||||||   :::::::::   ::::::::::::|
1013 eLysLysGlyIleAspGlyGluAspSerSerSerAspAlaThrSerAsnA 1030

3236 CACAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTT 3285
     ||:::|||:::   |||   :::      |||
1030 laAsnLeuThrIleLysThrLysGluLeuLys............LeuThr 1042

3286 GACGCGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGT 3335
     :::   |||      :::|||   |||         |||||||||   ::
1043 GluAspLeuSerIleSerGlyPheAsn.........LysAlaGluIleTh 1056

3336 TGCCGAACCGGCCCGGCAG......GCAGGCGGGGAAAATGCCGGCATTA 3379
     :|||:::   :::|||:::         |||:::::|||   |||
1056 rAlaLysAspGlyArgAspLeuThrIleGlyAsnSerAsnAspGlyAsnS 1073

3380 TGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCC 3429
        ||||||   :::           :::      ||||||:::
1073 erGlyAlaGluAlaLysThrValThrPheAsnAsnValLysAspSerLys 1089

3430 TTGGCGAAACAG..............CGCGAAGCGGAAACCCGGCCGGC 3464
     ::::::   :::               ::::::::::::::   ::
1090 IleSerAlaAspGlyHisAsnValThrLeuAsnSerLysValLysThrSe 1106

3465 TACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAAC 3514
     :::::::         |||      ::::::   :::|||
1106 rSerSerAsnGlyGlyArgGluSerAsnSerAspAsnAspThrGlyLeuT 1123

3515 CGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGC...CGT 3561
                ::::   :::   ::::::|||:::   |||   :::
1123 hrIleThrAlaLysAsnValGluValAsnLysAspIleThrSerLeuLys 1139

3562 TATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTT 3611
        :::|||         ||||||      :::||| :::::::::   :
1140 ThrValAsnIleThrAlaSerGluLysValThrThr.ThrAlaGlySerT 1156

3612 CGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACG 3661
     ::         ||||||      |||   :::         ::::::
1156 hrIleAsnAlaThrAsnGlyLysAlaSerIleThrThrLysThrGlyAsp 1172

3662 CCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGC...AA 3708
           |||      :::::::::|||   ::::::::::::
1173 IleSerGlyThrIleSerGlyAsnThrValSerValSerAlaThrGlyAs 1189

3709 GATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCA 3758
      :::::::::   :::::::::|||      :::   ::::::   ::::
1189 pLeuThrThrLysSerGlySerLysIleGluAlaLysSerGlyGluAlaA 1206

3759 GAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGA 3808
     ::   |||||||||:::|||:::   ::::::         |||   :::
1206 snValThrSerAlaThrGlyThrIleGlyGly.........ThrIleSer 1219

3809 CCGGAAACACCTTCGACGACGGCATCGGCA..................... 3838
        :::|||   ::::::||||||:::|||
1220 GlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeuThrValGlyAs 1236

3839 ................ACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGG 3872
                        :::   :::|||      |||:::      :::|
1236 nGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLeuThrAlaThrG 1253

3873 GCAAT......ACGGCATCGGCAGGTTCGACATCGGCATCAGCGCGGGCG 3916
     |||||      |||::::::|||||||||:::   :::|||:::::::::
1253 lyAsnThrLeuThrThrGluAlaGlySerSerIleThrSerThrLysGly 1269

3917 CGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGAGGCAAAATCCGC 3966
     :::|||   :::   |||      |||:::::::|||   |||   |||
1270 GlnValAspLeuLeuAla......GlnAsnGlySerIleAlaGlySerIl 1284

3967 CGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCGCGCAGGTTTCGG 4016
        |||:::   :::|||                |||      :::   |
1284 eAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrThrValA 1301

4017 CGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAA 4066
     ||   |||      :::   ||||||:::      :::|||::::::|||
1301 laGlySerAspIleLysAlaThrSerGlyThrLeuValIleAsnAlaLys 1317
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR63506

seq_documentation_block:

```
ID   AAR63506 standard; Protein; 1477 AA.
XX
AC   AAR63506;
XX
DT   25-JUN-1995  (first entry)
XX
DE   Haemophilus high molecular weight protein HMW2.
XX
KW   High molecular weight protein; HMW2; protective vaccine; otitis;
KW   sinusitis; bronchitis; Hib; ss.
XX
OS   Haemophilus.
XX
PN   WO9421290-A.
XX
PD   29-SEP-1994.
XX
PF   15-MAR-1994;   94WO-US02550.
XX
PR   16-MAR-1993;   93US-0038682.
XX
PA   (BARE/) BARENKAMP S J.
PA   (SGEM/) ST GEME J W.
XX
PI   Barenkamp SJ,  St GEME JW;
XX
DR   WPI; 1994-316665/39.
DR   Q-PSDB; Q72294.
XX
PT   New immunogenic high mol. wt. proteins of non typeable
PT   Haemophilus - useful in protective vaccines
XX
PS    Claim 3; Page 36; 127pp; English.
XX
CC   The HMW2 protein encoded by this sequence is useful in a vaccine to
CC   protect against disease caused by non-typeable Haemophilus which are
CC   not controlled by H. influenzae type b (Hib) vaccines.  The encoded
CC   protein can also be used as a carrier for protective Hib
CC   polysaccharide (in a conjugáte vaccin against meningitis) or for
CC   other antigens, haptens, etc.
XX
```

SQ Sequence 1477 AA;

alignment_scores:
Quality: 286.50 Length: 1325
Ratio: 0.448 Gaps: 62
Percent Similarity: 48.226 Percent Identity: 20.226

alignment_block:
US-09-303-518D-653 x AAR63506 ..

Align seg 1/1 to: AAR63506 from: 1 to: 1477

```
 706 AATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAAAATTAAACA 755
     ||||||      :::::::|||:::::::::::::::|||:::|||||| 
 165 AsnGlyPheThrAlaSerThrLeuAspIleSerAsnGluAsnIleLys.. 180

 756 TAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCT 805
                                                     :
 181 ..................................................A 181

 806 CACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGG 855
     ::      |||   ::::::      |||:::|||   |||   :::
 181 laArgAsnPheThrPheGluGlnThrLysAspLysAlaLeuAlaGlu... 196

 856 GTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCT 905
     :::::::::   |||         :::||||||:::::::::   :::||
 197 IleValAsnHisGlyLeuIleThrValGlyLysAspGlySerValAsnLe 213

 906 AGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATT 955
     |:::
 213 uIle..................................................... 214

 956 CAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAACGACAAT 1005
                           :::||||||
 215 .......................GlyGlyLysVal.............. 228

1006 AATAATGGCGCAGGAAAAATCGATGCCAAACATAAACACTATTCTCTACC 1055
        ::::::   |||   |||:::::::::            ||||||
 219 ...LysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeuLe 234

1056 TTATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCG 1105
               |||           :::::::::|||   :::::::::
 234 uAlaGlyGlnLysIleThrIleSerAspIleIleAsnProThrIleThrT 251

1106 AGACAGCAAGAGAACCT.........................GTTTAT 1128
       :::         |||                          ::::::
 251 yrSerIleAlaAlaProGluAsnGluAlaValAsnLeuGlyAspIlePhe 267

1129 CATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGA 1178
           :::|||:::::::|||                  ::::::|||
 268 AlaLysGlyGlyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGl 284

1179 AAATATTTCC......TTTATTGACAAAGGAAAA...GGTGAATTGATAC 1219
      :::::::|||         ::::::::|||   |||   |||:::::::::|
 284 yLysLeuSerAlaAspSerValSerLysAspLysSerGlyAsnIleValL 301

1220 TTACCAGCAACATCAACCAAGGC...GCGGGCGGTTTGTATTTTGAGGGT 1266
     ||:::::::::   :::::::::      ||||||:::
 301 euSerAlaLysGluGlyGluAlaGluIleGlyGlyVal.............. 313

1267 AATTTTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGT 1316
               :::|||   :::|||:::::::::   :::|||:::   ::
 314 .........IleSerAlaGlnAsnGlnGlnAla...LysGlyGlyLysLe 326

1317 TCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAA 1366
     :   |||:::   |||:::   ||||||   |||   :::|||:::::::
 326 uMetIleThr...GlyAspLysValThrLeuLys...ThrGlyAlaValI 341

1367 ACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTG...GTTCAAGCC 1413
       |||      :::|||   |||      |||   |||      :::
 341 leAspLeuSerGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAspGlu 357

1414 AAAGGGGAAAACCAAGGCTCGGTC.................AGCGTGGG 1445
     :::||||||:::::::::::::::                 |||:::
 358 ArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLysLysThrSerLeuGl 374

1446 CGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAAC 1495
        |||:::::|||      :::      :::      |||:::
 374 uLysGlySerThrIleAsnValSerGlyLysGluLysGlyGlyArgAlaI 391

1496 AAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTG 1545
       :::::::::::::|||:::|||::::::|||:::   ::::::|||
 391 leValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAlaGlnGly 407

1546 AATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGG 1595
     ::::::|||                                ::::::      ||
 408 SerGlyAsp.........................IleAlaLysThrGl 415

1596 CGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAA 1645
     ||||   ::::::   :::||||||   ||||||:::
 415 yGlyPheValGluThrSerGlyHisTyrLeuSerIle.............. 427

1646 ATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAA... 1692
          |||::::::|||::::::             :::   ||||||
 428 .....AspSerAsnAlaIleVal............LysThrLysGluTrp 438

1693 ............TCCACCGTTACCATT....................... 1707
                  :::   ||||||||
 439 LeuLeuAspProAspAspValThrIleGluAlaGluAspProLeuArgAs 455

1708 ....ACAGGCAATAAAGATATTACTACAACCGGCAATAACAACAACTTGG 1753
         ||||||   :::|||         ||||||:::::::::      |
 455 nAsnThrGlyIleAsnAspGluProProThrGlyThrGlyGluAlaSerA 472

1754 ATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCA 1803
     ||   |||||||::                        :::|||      :::
 472 spProLysLysAsn.....................SerGluLeuLysThr 481

1804 ACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAA........ 1845
     |||   ||||||      ::::::   ||||||      :::
 482 ThrLeuThrAsnThrThrIleSer...AsnTyrLeuLysAsnAlaTrpTh 497

1846 ..........GAAGCGGATCGCACTTTACTGCTTTCC.............. 1872
                |||:::|||   |||   ::::::
 497 rLeuAsnIleThrAlaSerArgLysLeuThrValAsnSerSerIleAsnI 514

1873 ..................................................GGC 1875
                                                       |||
 514 leGlySerAsnSerHisLeuIleLeuHisSerLysGlyGlnArgGlyGly 530

1876 GGAACAAATTTAAACGGCAATATCACGCAAACAAACGGCAAACTG..... 1920
     |||:::::::::::::|||:::||||||:::   :::|||:::|||
 531 GlyValGlnIleAspGlyAspIleThrSerLysGlyGlyAsnLeuThrIl 547

1921 .TTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCG 1969
      :::   :::|||            |||            |||   :::|
 547 eTyrSerGlyGlyTrpValAspValHisLysAsnIleThrLeuAspGlnG 564

1970 GGTGGTCAAAAATG.....................GAAGGT......... 1989
     ||:::   ::::::                     ||||||
 564 lyPheLeuAsnIleThrAlaAlaSerValAlaPheGluGlyGlyAsnAsn 580

1990 ............................ATCCCACAAGGAGAAATCGT 2009
                                  :::   ||||||   :::::
 581 LysAlaArgAspAlaAlaAsnAlaLysIleValAlaGlnGlyThrValTh 597

2010 GTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCCATA 2059
     :      ::::::      |||      |||:::|||:::|||      :
 597 rIleThrGlyGluGlyLysAsp......PheArgAlaAsnAsnValSerL 612

2060 TTCAGGGC.......................................... 2067
     :::::|||
 612 euAsnGlyThrGlyLysGlyLeuAsnIleIleSerServalAsnAsnLeu 628

2068 ............................GGACAAGCGGTGGTTTC 2084
```


629 ThrHisAsnLeuSerGlyThrIleAsnIleSerGlyAsnIleThrIleAs 645
2085 CCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCC 2134
645 nGlnThrThrArgLysAsnThrSerTyrTrpGlnThrSer..... 658
2135 AAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGT 2184
659 .....HisAspSerHis..... 662
2185 TCGGACTGG......ACGGGTCTGACAAGTTGTACCGAAAAAACCATTAC 2228
663 ......TrpAsnValSerAlaLeuAsnLeuGluThrGlyAlaAsnPheTh 677
2229 CGACGATAAAGTGATTGCTTCATTGAGCAAG.....ACCGACATCA 2269
677 rPheIleLysTyrIleSerSerAsnSerLysGlyLeuThrThrGlnTyrA 694
2270 GAGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTT 2319
694 rgSerSer.....AlaGlyValAsnPheAsnGly... 703
2320 GCCACACTCAACGGCAATCTTAGT.....GCAGGCGGAGACAC 2357
704 ......ValAsnGlyAsnMetSerPheAsnLeuLysGluGlyAlaLysVa 718
2358 GCACTATACGGTTACGCGCAACGCCACCCAAAAC..... 2391
718 lAsnPheLysLeuLysProAsnGluAsnMetAsnThrSerLysProLeuP 735
2392 .....GGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACA 2427
735 roIleArgPheLeuAlaAsnIleThrAlaThrGlyGlyGlySerValPhe 751
2428 TTTAAT...CAAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGC 2474
752 PheAspIleTyrAlaAsnHisSerGlyArgGlyAlaGluLeuLysMetSe 768
2475 TTCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTT 2524
768 rGluIleAsnIleSerAsnGlyAla.....AsnPheThrLeuA 781
2525 CCGACAACGCTAAGGCA.....AACGTAAGC 2550
781 snSerHisValArgAlaAspAspAlaPheLysIleAsnLysAspLeuThr 797
2551 CATTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCA 2600
798 IleAsnAlaThrAsnSerAsnPheSerLeuArgGlnThrLysAspAspPh 814
2601 TTTTGAA.....AACAGCCGCTTTACCGGAAAAA 2629
814 eTyrAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerI 831
2630 TCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACG 2679
831 leLeuGlyGlyAsnValThr...LeuGlyGlyGlnAsnSer..... 843
2680 CTGCCGTCGGGCACGGAATTAGGCAATTTAAACCTTGACAAC...GCCAC 2726
844 ......SerSerSerIleThrGlyAsnIleThrIleGluLysAlaAlaAs 858
2727 CATTACACTCAATTCCGCCTATCGACACGATGCGGCAGGCGCGCAAACCG 2776
858 nValThrLeu..... 861
2777 GCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCC 2826
862 ..GluAlaAsnAsnAlaProAsnGlnGlnAsnIleArgAspArgVal... 876
2827 GTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAAA 2876
877 .....IleLysLeuGlySerLeuLeuValAs 885
2877 CGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCG 2926
885 nGlySerLeuSerLeuThrGlyGluAsnAlaAspIle..... 897
2927 GCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCC..... 2964
898 .....LysGlyAsnLeuThrIleSerGluSerAlaThrPheLysGly 911
2965 .....GAAGGCACTTACACCTTGGCTGTCAA 2990
912 LysThrArgAspThrLeuAsnIleThrGlyAsnPheThr..... 924
2991 CAATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAG 3040
925 .AsnAsnGlyThrAlaGluIleAsnIleThrGlnGlyValValLysLeuG 941
3041 GAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAAC 3090
941 ly.....AsnValThrAsnAspGlyAspLeuAsnIleThrThrHisAla 955
3091 GAACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGG 3140
956 LysArgAsnGlnArgSerIleIleGlyGlyAspIleIleAsnLysLysGl 972
3141 CGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTT..... 3180
972 ySerLeuAsnIleThrAspSerAsnAsnAspAlaGluIleGlnIleGlyG 989
3181 .....TCCGACAAACTC 3192
989 lyAsnIleSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysIle 1005
3193 .....GGCAAGGCGGGAGAAACAGA 3212
1006 AsnIleThrLysGlnIleThrIleLysLysGlyIleAspGlyGluAspSe 1022
3213 GGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCGG 3262
1022 rSerSerAspAlaThrSerAsnAlaAsnLeuThrIleLysThrLysGluL 1039
3263 AAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCAAT 3312
1039 euLys.....LeuThrGluAspLeuSerIleSerGlyPheAsn 1051
3313 GCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAG.....GC 3356
1052 .....LysAlaGluIleThrAlaLysAspGlyArgAspLeuThrIl 1065
3357 AGGCGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGG 3406
1065 eGlyAsnSerAsnAspGlyAsnSerGlyAlaGluAlaLysThrValThrP 1082
3407 TGCAGGCGGATAAAGACACCGCCTTGGCGAAACAG..... 3441
1082 heAsnAsnValLysAspSerLysIleSerAlaAspGlyHisAsnValThr 1098
3442 CGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCG 3491
1099 LeuAsnSerLysValLysThrSerSerSerAsnGlyGlyArgGluSerAs 1115
3492 CGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGC 3541
1115 nSerAspAsnAspThrGlyLeuThrIleThrAlaLysAsnValGluValA 1132
3542 AGCGCGACCTGATCAGC...CGTTATGCCAATAGCGGTTTGAGTGAATTT 3588
1132 snLysAspIleThrSerLeuLysThrValAsnIleThrAlaSerGluLys 1148
3589 TCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGT 3638
1149 ValThrThr.ThrAlaGlySerThrIleAsnAlaThrAsnGlyLysAlaS 1165

```
     3639 GTTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACA 3688
          :::         ::::::        |||      ::::::::::|||
     1165 erIleThrThrLysThrGlyAspIleSerGlyThrIleSerGlyAsnThr 1181

     3689 CCAAACACTACCGTTCGC...AAGATTTCCGCGCCTACCGCCAACAAACC 3735
          ::::::::::::       :::::::::   :::::::::|||
     1182 ValSerValSerAlaThrGlyAspLeuThrThrLysSerGlySerLysIl 1198

     3736 GACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGG 3785
             :::   ::::::   ::::::   |||||||||:::|||:::   :
     1198 eGluAlaLysSerGlyGluAlaAsnValThrSerAlaThrGlyThrIleG 1215

     3786 CATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACGGCATCG 3835
          :::::         |||   :::   :::|||   ::::::||||||:::
     1215 lyGly.........ThrIleSerGlyAsnThrValAsnValThrAlaAsn 1228

     3836 GCA.....................................ACTCGGCACGG 3849
          |||                                    :::   :::||
     1229 AlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGl 1245

     3850 CTTGCCCACGGTGCCGTTTTCGGGCAAT......ACGGCATCGGCAGGTT 3893
          |      |||:::      :::|||||||      |||:::::|||||||
     1245 yAlaAlaThrLeuThrAlaThrGlyAsnThrLeuThrThrGluAlaGlyS 1262

     3894 CGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAG 3943
          ||:::   :::|||:::::::::::::|||   :::   |||     |||
     1262 erSerIleThrSerThrLysGlyGlnValAspLeuLeuAla......Gln 1276

     3944 ACGGCATCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAG 3993
          :::::::|||   |||   |||      |||:::   :::|||
     1277 AsnGlySerIleAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnTh 1293

     3994 GCAAGATACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGG 4043
               |||      :::   |||   |||      :::   ||||||:
     1293 rThrGlyThrLeuThrThrValAlaGlySerAspIleLysAlaThrSerG 1310

     4044 CGCAACGCGCTATTTCGTCCAAA 4066
          ::      :::|||:::::|||
     1310 lyThrLeuValIleAsnAlaLys 1317
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01825

seq_documentation_block:

```
ID   AAB01825 standard; Protein; 1221 AA.
XX
AC   AAB01825;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain Joyc mature HMW1A protein, SEQ ID NO:28.
XX
KW   Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain Joyc.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52176.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 8; Fig 18A-R; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   1221 AA;
```

alignment_scores:

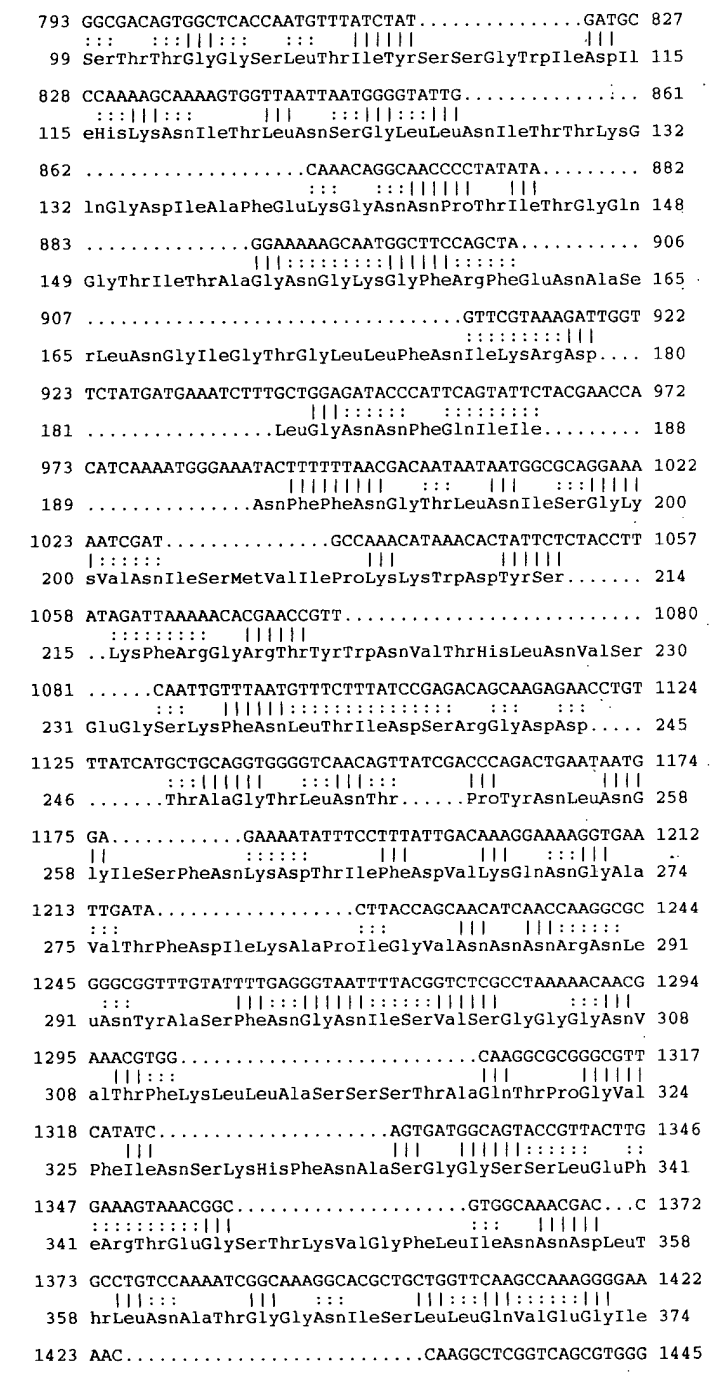
```
            Quality:  283.50              Length:   1217
              Ratio:  0.490                 Gaps:     62
 Percent Similarity:  47.576   Percent Identity: 20.542
```

alignment_block:

US-09-303-518D-653 x AAB01825 ..

Align seg 1/1 to: AAB01825 from: 1 to: 1221

```
     574 CCTGATCGTGTTCGAATCGGAGCAGGCAGACAATATTGGCGGTCTGAT.. 621
         ||||||:::|||   |||:::||||||   :::      |||:::|||
       1 ProAspAsnValSerIleAsnAlaGlyThrSerGlu...ArgAsnAspAl 16

     622 .........GAAGACGAACCC.................AATAACCGCG 643
                 ||||||   |||                  :::|||   :
      16 aSerProThrGluAspPheProThrGlyAlaGlyGlyLysAspAsnProL 33

     644 AAAGTTCATATCAT..............ATTGCAAGCGCATATTCTTGG 678
         ::::::::   |||              |||   ::::::   :::
      33 ysLysAsnAlaHisAsnLysProThrLeuIleAsnThrThrLeuGluArg 49

     679 CTCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGT 728
         :::::::::|||||||||||||:::   |||   ::::::         ::
      50 IleLeuSerGlyAsnThrPheVal...AsnIleThrAlaArgLysArgIl 65

     729 CAACTTAGGTAGCGAA...AAAATTAAACATAGCCCATAT.......... 765
         ::::::::::|||:::   :::||||||   |||   :::
      65 eThrValAsnSerAspIleAsnIleLysAspSerSerHisLeuIleLeuT 82

     766 ......................GGTTTTTTACCAACAGGAGGCTCATTT 792
                               |||                ||||:::
      82 rpSerGluAsnAspAsnSerSerGlyValAspIleLysGlyAsnIleThr 98
```


793 GGCGACAGTGGCTCACCAATGTTTATCTAT.....GATGC 827
99 SerThrThrGlyGlySerLeuThrIleTyrSerSerGlyTrpIleAspIl 115
828 CCAAAAGCAAAAGTGGTTAATTAATGGGGTATTG..... 861
115 eHisLysAsnIleThrLeuAsnSerGlyLeuLeuAsnIleThrThrLysG 132
862 .....CAAACAGGCAACCCCTATATA..... 882
132 lnGlyAspIleAlaPheGluLysGlyAsnAsnProThrIleThrGlyGln 148
883 .....GGAAAAAGCAATGGCTTCCAGCTA..... 906
149 GlyThrIleThrAlaGlyAsnGlyLysGlyPheArgPheGluAsnAlaSe 165
907 .....GTTCGTAAAGATTGGT 922
165 rLeuAsnGlyIleGlyThrGlyLeuLeuPheAsnIleLysArgAsp.... 180
923 TCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCA 972
181 .....LeuGlyAsnAsnPheGlnIleIle..... 188
973 CATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAA 1022
189 .....AsnPhePheAsnGlyThrLeuAsnIleSerGlyLy 200
1023 AATCGAT.....GCCAAACATAAACACTATTCTCTACCTT 1057
200 sValAsnIleSerMetValIleProLysLysTrpAspTyrSer..... 214
1058 ATAGATTAAAAACACGAACCGTT..... 1080
215 ..LysPheArgGlyArgThrTyrTrpAsnValThrHisLeuAsnValSer 230
1081 .....CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGT 1124
231 GluGlySerLysPheAsnLeuThrIleAspSerArgGlyAspAsp..... 245
1125 TTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATG 1174
246 .....ThrAlaGlyThrLeuAsnThr.....ProTyrAsnLeuAsnG 258
1175 GA.....GAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAA 1212
258 lyIleSerPheAsnLysAspThrIlePheAspValLysGlnAsnGlyAla 274
1213 TTGATA.....CTTACCAGCAACATCAACCAAGGCGC 1244
275 ValThrPheAspIleLysAlaProIleGlyValAsnAsnAsnArgAsnLe 291
1245 GGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCCTAAAAACAACG 1294
291 uAsnTyrAlaSerPheAsnGlyAsnIleSerValSerGlyGlyGlyAsnV 308
1295 AAACGTGG.....CAAGGCGCGGGCGTT 1317
308 alThrPheLysLeuLeuAlaSerSerSerThrAlaGlnThrProGlyVal 324
1318 CATATC.....AGTGATGGCAGTACCGTTACTTG 1346
325 PheIleAsnSerLysHisPheAsnAlaSerGlyGlySerSerLeuGluPh 341
1347 GAAAGTAAACGGC.....GTGGCAAACGAC...C 1372
341 eArgThrGluGlySerThrLysValGlyPheLeuIleAsnAsnAspLeuT 358
1373 GCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAA 1422
358 hrLeuAsnAlaThrGlyGlyAsnIleSerLeuLeuGlnValGluGlyIle 374
1423 AAC.....CAAGGCTCGGTCAGCGTGGG 1445
375 AspGlyMetIleGlyLysGlyValValAlaLysLysAsnIleThrPheAl 391
1446 CGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAAC 1495
391 aGlyGlyAsnIleThrPhe.....GlySerLysL 401
1496 AAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTG 1545
401 ysAlaIleThrGluIle.....GluGlyAsnAlaThrIleAsnAsn 414
1546 AATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGG 1595
415 AsnAlaAsn..... 417
1596 CGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAA 1645
418 .....ValThrLeuIleGlySerAspPheAspAsnHisGlnLysProL 432
1646 ATACCGATGAAGGGGCGATGATTGTCAACCACAATCAA.....GACAAA 1689
432 euThrIleLysLysAspValIleIleAsnSerGlyAsnLeuThrAlaGly 448
1690 GAATCCACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGCAA 1739
449 GlyAsnValIleAsnIleAsnGlyAsn.....LeuThrValAsnAs 462
1740 TAACAACAACTTGGATAGCAAAAAAGAAATTGCCTACAAC.....GGTT 1783
462 nGlyAlaAsnLeuLysAlaIleThrAsnPheThrPheAsnValGlyGlyL 479
1784 GGTTTGGCGAG..... 1794
479 euPheAspAsnLysGlyAsnSerAsnIleSerIleAlaArgGlyGlyAla 495
1795 .....AAAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCA 1838
496 LysPheLysAspIleAsnAsnThrSerSer...LeuAsnIleThrThrAs 511
1839 ACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAATTTAA 1888
511 nSerAspThrThrTyrArgThrIleIleGluGlyAsnIleThrAsnLysA 528
1889 ACGGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCG 1938
528 laGlyAspLeuAsnIleIleAspAsnLys..... 537
1939 ACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGG 1988
538 ...GlyAsnAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysGluGl 553
1989 TATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACAT 2038
553 y.....AsnLeuThrI 557
2039 TTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTTCCCGC 2088
557 leSerSerAspLysIleAsnIleThr...AsnGlnIleThrIleLysLys 572
2089 AATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAGC 2138
573 GlyValAsnLysGluAspSerAspSerSerThrAlaAsnAsnAlaAsnLe 589
2139 AGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGG 2188
589 u.....ThrIleLysThrLys.... 594
2189 ACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGACGATAAA 2238
595 .....GluLeuGlnLeuThrGlyAspLeu 602
2239 GTGATTGCTTCATTGAGCAAGACCGACATCAGA..... 2271

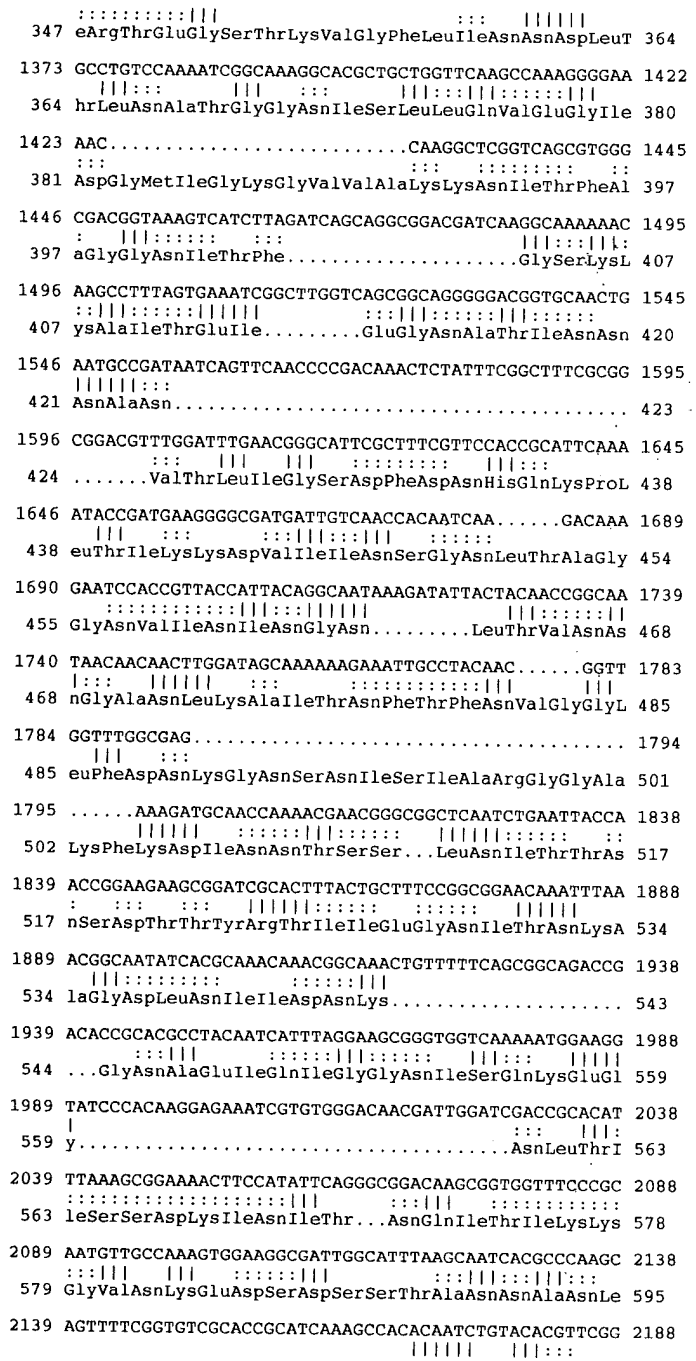
```
 603 AsnIleSerGlyPheAspLysAlaGluIleThrAlaLysGluGlyAlaAs 619

2272 .....GGCAAT.............................. 2277
            ||||||
 619 pLeuIleIleGlyAsnSerAspAsnAsnAsnAsnAlaAsnAlaLysLysV 636

2278 ........................GTCAGCCTTGCCGATCACGCT 2298
                            :::|||  ::::::|||
 636 alThrPheAsnGlnValLysAspSerLysIleSerAlaGlySerHisAsn 652

2299 CATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGCA.. 2346
               :::  :::  |||  |||||||||  :::|||
 653 ValThrLeuAsnSerLysValGluThrSerAsnGlyAsnAsnAspAlaGl 669

2347 .....GGCGGAGACACGCACTATACGGTT...ACGCGCAACGCCA 2383
           |||  :::|||     |||:::  ::::::|||:::|
 669 uSerAsnAsnGlyAspSerThrSerLeuThrIleAsnAlaLysAsnValT 686

2384 CCCAAAACGGCAACCTC.............................. 2400
     ||  |||:::|||:::
 686 hrValAsnAsnAsnIleThrSerHisLysThrValAsnIleThrAlaSer 702

2401 ................................AGCCTCGTGGGCAA 2414
                                    :::  :::|||::
 703 GluAsnValThrThrLysAlaGlyThrThrIleAsnAlaThrIleGlySe 719

2415 TGCCCAAGCAACATTTAATCAAGCCACATTAAACGGC............. 2451
     ::::::::::|||  :::  :::  ::::::|||
 719 rValGluValThrAlaLysThrGlyAspIleLysGlyGlyIleGluSerA 736

2452 ............AACACATCGGCTTCGGACAATGCTTCATTTAATCTA 2487
                |||  :::|||||||  :::  ::::::|||:::
 736 snSerGlyAsnValAsnIleThrAlaSerGlyAsp...ThrLeuAsnVal 751

2488 AGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGCTAA 2537
     ||||||  :::  ||||||  :::::::::  :::::::::|||
 752 SerAsnIleThrGlyGlnAsnValThrValAlaAlaAlaSerGlyAlaVa 768

2538 GGCAAACGTAAGCCATTCCGCACTCAAC.....GGCAATGTCTCCC 2578
      ::::::::::::  |||::::::|||       ||||||:::::::
 768 lThrThrThrLysGlySerThrIleAsnAlaThrThrGlyAsnAlaAsnI 785

2579 TAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGAAAA 2628
     :::::  |||                          |||||||:::
 785 leThrThrLys.............................ThrGlyGlu 791

2629 ATCAGCGGCGGCAAGGATACGGCA.....TTACACTTAAAAGACAG 2669
     |||:::|||         :::|||         :::::::::      ||
 792 IleAsnGlyGluValLysSerAlaSerGlyAsnValAsnIleThrAlaSe 808

2670 CGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAACCTTGACA 2719
     |      ||||||         :::::::::::|||::::::   :::|
 808 rGlyAsnThrLeu.....AsnAsnValSerAsnIleThrGlyGlnA 822

2720 ACGCCACCATTACACTCAATTCCGCCTATCGACACGATGCGGCAGGCGCG 2769
     ||:::|||:::|||  ||||||:::            |||      :::
 822 snValThrValThrAlaAsnSerGly.....AlaIleThrThr 834

2770 CAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCT 2819
           ||||||:::  :::|||
 835 ThrGluGlySerThrIleAsnAla..................... 842

2820 ATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGA 2869
                    |||:::::::::::::::::::::::|||  |
 843 .................ThrThrGlyAspAlaAsnIleThrThrGlnT 853

2870 CG......GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATG 2913
     ||      :::|||||||||:::::::::::
 853 hrGlyAsnIleAsnGlyLysValGluSerSer.............. 863

2914 TCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTC 2963
                       ||||||::::::  |||     :::::
 864 ..................SerGlySerValThrLeuIleAlaThrGl 873

2964 CGAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAA 3013
     ::::        ||||||||||||:::|||  :::::::::  |||:
 873 yGln.....ThrLeuAlaValGlyAsnIleSerGlyAspThrValT 887

3014 GTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCC 3063
     ::        :::|||:::  :::||||||  :::|||     |||
 887 hr.....IleThrAlaAspLysGlyLysLeuThrThrGlnThrSer 900

3064 GAAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCG.... 3109
     :::::::::|||  |||  ::::::::::||||||     :::
 901 SerLysIleAsnGlyThrLys.SerValThrThrSerSerGlnSerGlyA 917

3110 .....CATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGC 3154
          |||:::||||||  ::::::  ::::::|||::::::
 917 spIleSerGlyThrIleSerGlyAsnThrValSerValSerAlaThrGly 933

3155 ATAATCCGGTCAAAGAA....CAAGAGCTTTCCGACAAACTCGGCAAGGC 3200
        :::  ::::::     ::::::::::::  |||  |||:::||
 934 SerLeuThrThrGlnAlaGlySerLysIleGluAlaLysThrGlyGluAl 950

3201 GGGAGAAACAGAGGCCGCC.............................. 3219
     |:::  |||:::|||:::
 950 aAsnValThrSerAlaThrGlyThrIleGlyGlyThrIleSerGlyAsnT 967

3220 ........TTGACGGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCG 3261
          :::|||||||:::     :::|||:::  |||:::  |||
 967 hrValAsnValThrAlaAsnThrAspAsnLeuThrIleLysAspGlyAla 983

3262 GAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCAA 3311
     :::      :::  :::     :::|||  |||:::|||  ::
 984 ArgIleLysAlaThrGlyGlyAlaValThrLeuThrAlaThrGlyGlyTh 1000

3312 TGCCACCGAAAAGGCAGAAAGT.......................... 3333
     :  |||  :::::::::|||
1000 rLeuThrThrGluThrSerSerAspIleThrSerSerAsnGlyGlnThrT 1017

3334 .....GTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGCC..... 3372
          :::|||:::  :::     |||||||:::  ||||||
1017 hrLeuThrAlaLysAspSerSerIleAlaGlySerIleAsnAlaAlaAsn 1033

3373 .................GGCATTATGCAGGCGGAGGAAGAGAAAAAACG 3404
                       |||  :::  :::        :::|||
1034 ValThrLeuAsnThrThrGlyThrLeuThrThrValAlaGlySerLys.. 1049

3405 GGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAA 3453
      :::::::|||  :::  |||  :::  :::  ::::::|||:::
1050 .IleGluAlaAlaSerGlyThrLeuValIleAsnAlaLysAspAlaGln 1065
```

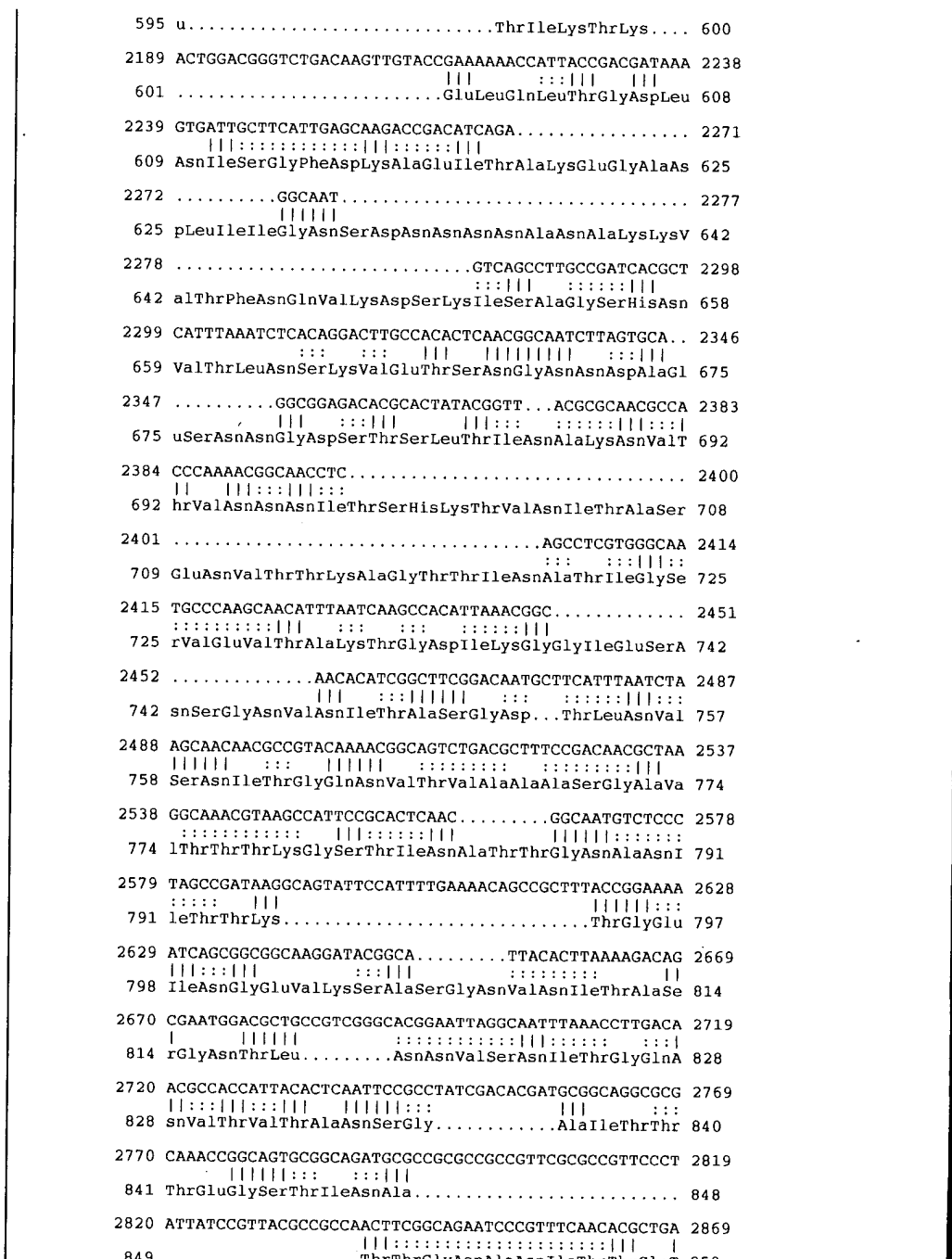
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01824

seq_documentation_block:

ID AAB01824 standard; Protein; 1227 AA.
XX
AC AAB01824;
XX
DT 11-SEP-2000 (first entry).
XX
DE Haemophilus influenzae strain Joyc HMW1A protein, SEQ ID NO:26.
XX
KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
OS Haemophilus influenzae strain Joyc.
XX
PN WO200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.

```
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52175.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 18A-R; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1227 AA;

alignment_scores:
            Quality:  283.50              Length:   1217
              Ratio:  0.490                Gaps:     62
 Percent Similarity:  47.576   Percent Identity: 20.542

alignment_block:
 US-09-303-518D-653 x AAB01824    ..

 Align seg 1/1  to: AAB01824  from: 1  to: 1227

     574 CCTGATCGTGTTCGAATCGGAGCAGGCAGACAATATTGGCGGTCTGAT.. 621
         ||||||:::|||   |||:::||||||   :::      |||:::|||
       7 ProAspAsnValSerIleAsnAlaGlyThrSerGlu...ArgAsnAspAl 22

     622 ..........GAAGACGAACCC..................AATAACCGCG 643
                   ||||||   |||                  :::|||   :
      22 aSerProThrGluAspPheProThrGlyAlaGlyGlyLysAspAsnProL 39

     644 AAAGTTCATATCAT..............ATTGCAAGCGCATATTCTTGG 678
         ::::::::   |||              |||   ::::::   :::
      39 ysLysAsnAlaHisAsnLysProThrLeuIleAsnThrThrLeuGluArg 55

     679 CTCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGT 728
         :::::::::|||||||||||||:::   |||   ::::::      ::
      56 IleLeuSerGlyAsnThrPheVal...AsnIleThrAlaArgLysArgIl 71

     729 CAACTTAGGTAGCGAA...AAAATTAAACATAGCCCATAT......... 765
         ::::::::::|||:::   :::||||||   |||   :::
      71 eThrValAsnSerAspIleAsnIleLysAspSerSerHisLeuIleLeuT 88

     766 ......................GGTTTTTTACCAACAGGAGGCTCATTT 792
                                |||              |||:::
      88 rpSerGluAsnAspAsnSerSerGlyValAspIleLysGlyAsnIleThr 104

     793 GGCGACAGTGGCTCACCAATGTTTATCTAT..............GATGC 827
         :::   :::|||:::   :::   ||||||              |||
     105 SerThrThrGlyGlySerLeuThrIleTyrSerSerGlyTrpIleAspIl 121

     828 CCAAAAGCAAAAGTGGTTAATTAATGGGGTATTG............... 861
          :::|||:::      |||   :::|||:::|||
     121 eHisLysAsnIleThrLeuAsnSerGlyLeuLeuAsnIleThrThrLysG 138

     862 ....................CAAACAGGCAACCCCTATATA......... 882
                              :::   :::||||||   |||
     138 lnGlyAspIleAlaPheGluLysGlyAsnAsnProThrIleThrGlyGln 154

     883 ..............GGAAAAAGCAATGGCTTCCAGCTA........... 906
                     |||::::::::::||||||::::::
     155 GlyThrIleThrAlaGlyAsnGlyLysGlyPheArgPheGluAsnAlaSe 171

     907 ................................GTTCGTAAAGATTGGT 922
                                        :::::::::|||
     171 rLeuAsnGlyIleGlyThrGlyLeuLeuPheAsnIleLysArgAsp.... 186

     923 TCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCA 972
                            |||:::::::   :::::::::
     187 .................LeuGlyAsnAsnPheGlnIleIle......... 194

     973 CATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAA 1022
                          |||||||||   :::   |||   :::||||||
     195 ..............AsnPhePheAsnGlyThrLeuAsnIleSerGlyLy 206

    1023 AATCGAT..............GCCAAACATAAACACTATTCTCTACCTT 1057
         |:::::::                |||          ||||||
     206 sValAsnIleSerMetValIleProLysLysTrpAspTyrSer....... 220

    1058 ATAGATTAAAAACACGAACCGTT.......................... 1080
           :::::::::   ||||||
     221 ..LysPheArgGlyArgThrTyrTrpAsnValThrHisLeuAsnValSer 236

    1081 ......CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGT 1124
                :::   ||||||:::::::::::::::   :::   :::
     237 GluGlySerLysPheAsnLeuThrIleAspSerArgGlyAspAsp..... 251

    1125 TTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATG 1174
                ::::||||||   :::|||:::      |||         ||||
     252 .......ThrAlaGlyThrLeuAsnThr......ProTyrAsnLeuAsnG 264

    1175 GA............GAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAA 1212
         ||            ::::::      |||      |||   ::::||
     264 lyIleSerPheAsnLysAspThrIlePheAspValLysGlnAsnGlyAla 280

    1213 TTGATA.................CTTACCAGCAACATCAACCAAGGCGC 1244
         :::                     :::      |||   ||||::::::
     281 ValThrPheAspIleLysAlaProIleGlyValAsnAsnAsnArgAsnLe 297

    1245 GGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCCTAAAAACAACG 1294
          :::         |||:::||||||:::::||||||      ::::||
     297 uAsnTyrAlaSerPheAsnGlyAsnIleSerValSerGlyGlyGlyAsnV 314

    1295 AAACGTGG.........................CAAGGCGCGGGCGTT 1317
           ||:::                          |||      ||||||
     314 alThrPheLysLeuLeuAlaSerSerSerThrAlaGlnThrProGlyVal 330

    1318 CATATC....................AGTGATGGCAGTACCGTTACTTG 1346
           |||                     |||   ||||||::::::   ::
     331 PheIleAsnSerLysHisPheAsnAlaSerGlyGlySerSerLeuGluPh 347

    1347 GAAAGTAAACGGC....................GTGGCAAACGAC...C 1372
```


347 eArgThrGluGlySerThrLysValGlyPheLeuIleAsnAsnAspLeuT 364
1373 GCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAA 1422
364 hrLeuAsnAlaThrGlyGlyAsnIleSerLeuLeuGlnValGluGlyIle 380
1423 AAC.....CAAGGCTCGGTCAGCGTGGG 1445
381 AspGlyMetIleGlyLysGlyValValAlaLysLysAsnIleThrPheAl 397
1446 CGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAAC 1495
397 aGlyGlyAsnIleThrPhe.....GlySerLysL 407
1496 AAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTG 1545
407 ysAlaIleThrGluIle.....GluGlyAsnAlaThrIleAsnAsn 420
1546 AATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGG 1595
421 AsnAlaAsn..... 423
1596 CGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAA 1645
424 .....ValThrLeuIleGlySerAspPheAspAsnHisGlnLysProL 438
1646 ATACCGATGAAGGGGCGATGATTGTCAACCACAATCAA.....GACAAA 1689
438 euThrIleLysLysAspValIleIleAsnSerGlyAsnLeuThrAlaGly 454
1690 GAATCCACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGCAA 1739
455 GlyAsnValIleAsnIleAsnGlyAsn.....LeuThrValAsnAs 468
1740 TAACAACAACTTGGATAGCAAAAAAGAAATTGCCTACAAC.....GGTT 1783
468 nGlyAlaAsnLeuLysAlaIleThrAsnPheThrPheAsnValGlyGlyL 485
1784 GGTTTGGCGAG..... 1794
485 euPheAspAsnLysGlyAsnSerAsnIleSerIleAlaArgGlyGlyAla 501
1795 .....AAAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCA 1838
502 LysPheLysAspIleAsnAsnThrSerSer...LeuAsnIleThrThrAs 517
1839 ACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAATTTAA 1888
517 nSerAspThrThrTyrArgThrIleIleGluGlyAsnIleThrAsnLysA 534
1889 ACGGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCG 1938
534 laGlyAspLeuAsnIleIleAspAsnLys..... 543
1939 ACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGG 1988
544 ...GlyAsnAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysGluGl 559
1989 TATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACAT 2038
559 y.....AsnLeuThrI 563
2039 TTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTTCCCGC 2088
563 leSerSerAspLysIleAsnIleThr...AsnGlnIleThrIleLysLys 578
2089 AATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAGC 2138
579 GlyValAsnLysGluAspSerAspSerSerThrAlaAsnAsnAlaAsnLe 595
2139 AGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGG 2188


595 u.....ThrIleLysThrLys.... 600
2189 ACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGACGATAAA 2238
601 .....GluLeuGlnLeuThrGlyAspLeu 608
2239 GTGATTGCTTCATTGAGCAAGACCGACATCAGA..... 2271
609 AsnIleSerGlyPheAspLysAlaGluIleThrAlaLysGluGlyAlaAs 625
2272 .....GGCAAT..... 2277
625 pLeuIleIleGlyAsnSerAspAsnAsnAsnAsnAlaAsnAlaLysLysV 642
2278 .....GTCAGCCTTGCCGATCACGCT 2298
642 alThrPheAsnGlnValLysAspSerLysIleSerAlaGlySerHisAsn 658
2299 CATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGCA.. 2346
659 ValThrLeuAsnSerLysValGluThrSerAsnGlyAsnAsnAspAlaGl 675
2347 .....GGCGGAGACACGCACTATACGGTT...ACGCGCAACGCCA 2383
675 uSerAsnAsnGlyAspSerThrSerLeuThrIleAsnAlaLysAsnValT 692
2384 CCCAAAACGGCAACCTC..... 2400
692 hrValAsnAsnAsnIleThrSerHisLysThrValAsnIleThrAlaSer 708
2401 .....AGCCTCGTGGGCAA 2414
709 GluAsnValThrThrLysAlaGlyThrThrIleAsnAlaThrIleGlySe 725
2415 TGCCCAAGCAACATTTAATCAAGCCACATTAAACGGC..... 2451
725 rValGluValThrAlaLysThrGlyAspIleLysGlyGlyIleGluSerA 742
2452 .....AACACATCGGCTTCGGACAATGCTTCATTTAATCTA 2487
742 snSerGlyAsnValAsnIleThrAlaSerGlyAsp...ThrLeuAsnVal 757
2488 AGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGCTAA 2537
758 SerAsnIleThrGlyGlnAsnValThrValAlaAlaAlaSerGlyAlaVa 774
2538 GGCAAACGTAAGCCATTCCGCACTCAAC.....GGCAATGTCTCCC 2578
774 lThrThrThrLysGlySerThrIleAsnAlaThrThrGlyAsnAlaAsnI 791
2579 TAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGAAAA 2628
791 leThrThrLys.....ThrGlyGlu 797
2629 ATCAGCGGCGGCAAGGATACGGCA.....TTACACTTAAAAGACAG 2669
798 IleAsnGlyGluValLysSerAlaSerGlyAsnValAsnIleThrAlaSe 814
2670 CGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAACCTTGACA 2719
814 rGlyAsnThrLeu.....AsnAsnValSerAsnIleThrGlyGlnA 828
2720 ACGCCACCATTACACTCAATTCCGCCTATCGACACGATGCGGCAGGCGCG 2769
828 snValThrValThrAlaAsnSerGly.....AlaIleThrThr 840
2770 CAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCT 2819
841 ThrGluGlySerThrIleAsnAla..... 848
2820 ATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGA 2869
849 .....ThrThrGlyAspAlaAsnIleThrThrGlnT 859

```
2870 CG......GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATG 2913
     ||      :::||||||||||:::::::::::
 859 hrGlyAsnIleAsnGlyLysValGluSerSer..................  869

2914 TCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTC 2963
                          ||||||::::::   |||      :::::
 870 .....................SerGlySerValThrLeuIleAlaThrGl  879

2964 CGAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAA 3013
     ::::         ||||||||||||:::|||   :::::::::   |||:
 879 yGln.........ThrLeuAlaValGlyAsnIleSerGlyAspThrValT  893

3014 GTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCC 3063
     ::         :::|||:::   :::|||||||   :::|||      |||
 893 hr.........IleThrAlaAspLysGlyLysLeuThrThrGlnThrSer  906

3064 GAAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCG.... 3109
     :::::::::|||   |||    :::::::::||||||      :::
 907 SerLysIleAsnGlyThrLys.SerValThrThrSerSerGlnSerGlyA  923

3110 .....CATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGC 3154
           |||:::||||||   ::::::   ::::::|||::::::
 923 spIleSerGlyThrIleSerGlyAsnThrValSerValSerAlaThrGly  939

3155 ATAATCCGGTCAAAGAA....CAAGAGCTTTCCGACAAACTCGGCAAGGC 3200
        :::   ::::::      ::::::::::::   |||   |||:::||
 940 SerLeuThrThrGlnAlaGlySerLysIleGluAlaLysThrGlyGluAl  956

3201 GGGAGAAACAGAGGCCGCC.............................. 3219
     |:::   |||:::|||:::
 956 aAsnValThrSerAlaThrGlyThrIleGlyGlyThrIleSerGlyAsnT  973

3220 ........TTGACGGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCG 3261
             :::||||||:::      :::|||:::   |||:::   |||
 973 hrValAsnValThrAlaAsnThrAspAsnLeuThrIleLysAspGlyAla  989

3262 GAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCAA 3311
     :::         :::   :::      :::|||   |||:::|||   ::
 990 ArgIleLysAlaThrGlyGlyAlaValThrLeuThrAlaThrGlyGlyTh 1006

3312 TGCCACCGAAAAGGCAGAAAGT........................... 3333
     :   |||   :::::::::|||
1006 rLeuThrThrGluThrSerSerAspIleThrSerSerAsnGlyGlnThrT 1023

3334 .....GTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGCC...... 3372
          :::|||:::   :::      ||||||:::   ||||||
1023 hrLeuThrAlaLysAspSerSerIleAlaGlySerIleAsnAlaAlaAsn 1039

3373 .................GGCATTATGCAGGCGGAGGAAGAGAAAAAACG 3404
                       |||   :::   :::         :::|||
1040 ValThrLeuAsnThrThrGlyThrLeuThrThrValAlaGlySerLys.. 1055

3405 GGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAA 3453
      ::::::|||   :::   |||   :::   :::   ::::::|||:::
1056 .IleGluAlaAlaSerGlyThrLeuValIleAsnAlaLysAspAlaGln 1071
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01848

seq_documentation_block:

```
ID   AAB01848 standard; Protein; 1477 AA.
XX
AC   AAB01848;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain 12 HMW2A protein, SEQ ID NO:71.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain 12.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52197.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Example 16; Fig 29A-N; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1477 AA;
```

```
alignment_scores:
            Quality:  282.50              Length:   1322
              Ratio:  0.440                 Gaps:     62
 Percent Similarity:  48.563   Percent Identity: 20.045

alignment_block:
 US-09-303-518D-653 x AAB01848   ..

 Align seg 1/1  to: AAB01848  from: 1  to: 1477

     706 AATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAAAATTAAACA 755
         ||||||      :::::::|||:::::::::::::::|||:::||||||
     165 AsnGlyPheThrAlaSerThrLeuAspIleSerAsnGluAsnIleLys.. 180

     756 TAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCT 805
                                                         :
     181 .............................................A 181
```

```
 806 CACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGG 855
     ::      |||   ::::::      |||:::|||   |||   :::
 181 laArgAsnPheThrPheGluGlnThrLysAspLysAlaLeuAlaGlu... 196

 856 GTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCT 905
     :::::::::   |||         :::||||||:::::::::   :::||
 197 IleValAsnHisGlyLeuIleThrValGlyLysAspGlySerValAsnLe 213

 906 AGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATT 955
     |:::
 213 uIle..................................................... 214

 956 CAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAACGACAAT 1005
                           :::||||||
 215 .......................GlyGlyLysVal.............. 218

1006 AATAATGGCGCAGGAAAAATCGATGCCAAACATAAACACTATTCTCTACC 1055
        ::::::   |||   |||:::::::::             ||||||
 219 ...LysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeuLe 234

1056 TTATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCG 1105
               |||              ::::::::::|||   :::::::::
 234 uAlaGlyGlnLysIleThrIleSerAspIleIleAsnProThrIleThrT 251

1106 AGACAGCAAGAGAACCT.........................GTTTAT 1128
     :::          |||                          ::::::
 251 yrSerIleAlaAlaProGluAsnGluAlaValAsnLeuGlyAspIlePhe 267

1129 CATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGA 1178
           :::|||:::::::|||                :::::::|||
 268 AlaLysGlyGlyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGl 284

1179 AAATATTTCC......TTTATTGACAAAGGAAAA...GGTGAATTGATAC 1219
      :::::::|||         ::::::|||   |||   |||::::::::::|
 284 yLysLeuSerAlaAspSerValSerLysAspLysSerGlyAsnIleValL 301

1220 TTACCAGCAACATCAACCAAGGC...GCGGGCGGTTTGTATTTTGAGGGT 1266
     ||::::::::::   :::::::::      |||||||:::
 301 euSerAlaLysGluGlyGluAlaGluIleGlyGlyVal.............. 313

1267 AATTTTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGT 1316
              :::|||   :::|||::::::::::   :::|||:::  :::
 314 .........IleSerAlaGlnAsnGlnGlnAla...LysGlyGlyLysLe 326

1317 TCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAA 1366
     :   |||:::   |||:::   ||||||   |||   :::|||::::::
 326 uMetIleThr...GlyAspLysValThrLeuLys...ThrGlyAlaVal I 341

1367 ACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTG...GTTCAAGCC 1413
       |||       :::|||   |||       |||   |||    :::
 341 leAspLeuSerGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAspGlu 357

1414 AAAGGGGAAAACCAAGGCTCGGTC.................AGCGTGGG 1445
     :::||||||:::::::::::::::                 |||:::
 358 ArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLysLysThrSerLeuGl 374

1446 CGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAAC 1495
        |||::::::|||      :::      :::      |||:::
 374 uLysGlySerThrIleAsnValSerGlyLysGluLysGlyGlyArgAlaI 391

1496 AAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTG 1545
       ::::::::::::|||:::|||::::::|||:::   ::::::|||
 391 leValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAlaGlnGly 407

1546 AATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGG 1595
     ::::::|||                          ::::::      ||
 408 SerGlyAsp.........................IleAlaLysThrGl 415

1596 CGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAA 1645
     ||||   ::::::   :::||||||   |||||||:::
 415 yGlyPheValGluThrSerGlyHisTyrLeuSerIle.............. 427

1646 ATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAA... 1692
```

```
          |||:::::::|||::::::              :::   ||||||
 428 .....AspSerAsnAlaIleVal............LysThrLysGluTrp 438

1693 ............TCCACCGTTACCATT..................... 1707
                 :::   |||||||||
 439 LeuLeuAspProAspAspValThrIleGluAlaGluAspProLeuArgAs 455

1708 ....ACAGGCAATAAAGATATTACTACAACCGGCAATAACAACAACTTGG 1753
         ||||||   :::|||          ||||||:::::::::      |
 455 nAsnThrGlyIleAsnAspGluPheProThrGlyThrGlyGluAlaSerA 472

1754 ATAGCAAAAAA..................................... 1764
     ||   ||||||
 472 spProLysLysAsnSerGluLeuLysThrThrLeuThrAsnThrThrIle 488

1765 ...GAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAA.. 1809
        :::          |||:::|||     :::      |||::::::
 489 SerAsnTyrLeuLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLy 505

1810 .......ACGAACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAGCGG 1852
            :::|||:::   :::|||::::::               ::::
 505 sLeuThrValAsnSerSerIleAsnIleGly..............SerA 517

1853 ATCGCACTTTACTGCTT..................TCCGGCGGAACAAAT 1884
     ::      |||:::|||                  :::||||||::::::
 517 snSerHisLeuIleLeuHisSerLysGlyGlnArgGlyGlyGlyValGln 533

1885 TTAAACGGCAATATCACGCAAACAAACGGCAAACTG......TTTTTCAG 1928
     ::::::|||:::||||||:::   :::|||:::|||      :::   ::
 534 IleAspGlyAspIleThrSerLysGlyGlyAsnLeuThrIleTyrSerGl 550

1929 CGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAA 1978
     :|||          |||          |||   :::|||:::   :
 550 yGlyTrpValAspValHisLysAsnIleThrLeuAspGlnGlyPheLeuA 567

1979 AAATG.....................GAAGGT.................. 1989
     :::::                     ||||||
 567 snIleThrAlaAlaSerValAlaPheGluGlyGlyAsnAsnLysAlaArg 583

1990 .....................ATCCCACAAGGAGAAATCGTGTGGGACAA 2018
                          :::   ||||||   ::::::      ::
 584 AspAlaAlaAsnAlaLysIleValAlaGlnGlyThrValThrIleThrGl 600

2019 CGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCCATATTCAGGGC. 2067
     ::::       |||       |||:::|||:::|||       :::::|||
 600 yGluGlyLysAsp......PheArgAlaAsnAsnValSerLeuAsnGlyT 615

2067 ..................................................... 2067

 615 hrGlyLysGlyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsn 631

2068 .......................GGACAAGCGGTGGTTTCCCGCAATGT 2093
                            |||:::   :::::::::::::::::
 632 LeuSerGlyThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThrTh 648

2094 TGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTT 2143
     :   |||       :::   |||:::   |||
 648 rArgLysAsnThrSerTyrTrpGlnThrSer..................... 658

2144 TCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGG 2193
                  |||:::||||||                       |||
 659 .............HisAspSerHis.....................Trp 663

2194 ......ACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGACGATAA 2237
           :::::::|||:::      |||      :::::::|||      ||
 664 AsnValSerAlaLeuAsnLeuGluThrGlyAlaAsnPheThrPheIleLy 680

2238 AGTGATTGCTTCATTGAGCAAG.........ACCGACATCAGAGGCAATG 2278
     |   |||:::|||   ||||||         |||:::   |||::::::
 680 sTyrIleSerSerAsnSerLysGlyLeuThrThrGlnTyrArgSerSer. 696

2279 TCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACACTC 2328
                        |||   :::|||::::::|||         :::
```

```
 697 .................AlaGlyValAsnPheAsnGly.........Val 704

2329 AACGGCAATCTTAGT............GCAGGCGGAGACACGCACTATAC 2366
     |||||||||:::|||               |||:::   :::::::::
 705 AsnGlyAsnMetSerPheAsnLeuLysGluGlyAlaLysValAsnPheLy 721

2367 GGTTACGCGCAACGCCACCCAAAAC........................ 2391
      :::      |||   :::::::|||
 721 sLeuLysProAsnGluAsnMetAsnThrSerLysProLeuProIleArgP 738

2392 .....GGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAAT... 2433
          :::|||:::::::   :::|||:::::::::::::   |||:::
 738 heLeuAlaAsnIleThrAlaThrGlyGlyGlySerValPhePheAspIle 754

2434 CAAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAA 2483
        |||:::   :::|||:::   :::           :::::::::||
 755 TyrAlaAsnHisSerGlyArgGlyAlaGluLeuLysMetSerGluIleAs 771

2484 TCTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACG 2533
     |:::||||||:::|||           ::::::||||||:::::::::::
 771 nIleSerAsnGlyAla............AsnPheThrLeuAsnSerHisV 784

2534 CTAAGGCA.......................AACGTAAGCCATTCCGCA 2559
     ::::::::                       :::::::::   :::|||
 784 alArgGlyAspAspAlaPheLysIleAsnLysAspLeuThrIleAsnAla 800

2560 CTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAA.. 2607
        |||:::|||   ||||||   :::                 ::::::
 801 ThrAsnSerAsnPheSerLeuArgGlnThrLysAspAspPheTyrAspGl 817

2608 ..................AACAGCCGCTTTACCGGAAAAATCAGCGGCG 2638
                        ||||||   ::::::   :::|||   ||||
 817 yTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIleLeuGlyG 834

2639 GCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCG 2688
     ||:::   |||   |||      ::::::|||                |||
 834 lyAsnValThr...LeuGlyGlyGlnAsnSer..............Ser 844

2689 GGCACGGAATTAGGCAATTTAAACCTTGACAAC...GCCACCATTACACT 2735
     ::::::      ||||||:::::::::::::::   |||::::::|||||
 845 SerSerIleThrGlyAsnIleThrIleGluLysAlaAlaAsnValThrLe 861

2736 CAATTCCGCCTATCGACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGG 2785
     |                                          :::|||
 861 u.........................................GluAlaA 864

2786 CAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCG 2835
       :::||||||:::::::::::   |||:::   :::
 864 snAsnAlaProAsnGlnGlnAsnIleArgAspArgVal..... 876

2836 CCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATT 2885
                          :::::::::::|||   |||||||||:::||
 877 ..............IleLysLeuGlySerLeuLeuValAsnGlySerLe 888

2886 GAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCA 2935
     |:::      |||            :::                 :
 888 uSerLeuThrGlyGluAsnAlaAspIle.....................L 898

2936 GCGGCAAATTGAAGCTGGCGGAAAGTTCC..................... 2964
     ::|||:::|||   ::::::||||||:::
 898 ysGlyAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArg 914

2965 ..............GAAGGCACTTACACCTTGGCTGTCAACAATACCGG 2999
                       |||::::::|||            |||:::||
 915 AspThrLeuAsnIleThrGlyAsnPheThr............AsnAsnGl 927

3000 CAACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACA 3049
     }:::      :::::::::   |||   :::|||      |||      |
 927 yThrAlaGluIleAsnIleThrGlnGlyValValLysLeuGly......A 942

3050 ACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTC 3099
     ||:::      :::   :::||||||:::|||   :::   ::::::
 942 snValThrAsnAspGlyAspLeuAsnIleThrThrHisAlaLysArgAsn 958

3100 GATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCG 3149
     :::   :::            ::::::|||:::|||   |||::::::::
 959 GlnArgSerIleIleGlyGlyAspIleIleAsnLysLysGlySerLeuAs 975

3150 CCTGCATAATCCGGTCAAAGAACAAGAGCTT.................. 3180
     ::::   :::      ::::::   |||:::
 975 nIleThrAspSerAsnAsnAspAlaGluIleGlnIleGlyGlyAsnIleS 992

3181 ............................TCCGACAAACTC......... 3192
                                 |||||||||:::
 992 erGlnLysGluGlyAsnLeuThrIleSerSerAspLysIleAsnIleThr 1008

3193 .....................GGCAAGGCGGGAGAAACAGAGGCCGCCTT 3221
                          |||      ||||||   ::::::::
1009 LysGlnIleThrIleLysLysGlyIleAspGlyGluAspSerSerSerAs 1025

3222 GACGGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCGGAAAAAGACA 3271
      ::::::::::::|||:::|||:::   |||   :::      |||
1025 pAlaThrSerAsnAlaAsnLeuThrIleLysThrLysGluLeuLys.... 1040

3272 ACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCAATGCCACCGAA 3321
                    :::   |||      :::|||   |||
1041 ........LeuThrGluAspLeuSerIleSerGlyPheAsn......... 1051

3322 AAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAG......GCAGGCGGGGA 3365
     |||||||||   :::|||:::   :::|||:::         |||:::::
1052 LysAlaGluIleThrAlaLysAspGlyArgAspLeuThrIleGlyAsnSe 1068

3366 AAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGG 3415
     :|||   |||         ||||||   :::            :::
1068 rAsnAspGlyAsnSerGlyAlaGluAlaLysThrValThrPheAsnAsnV 1085

3416 ATAAAGACACCGCCTTGGCGAAACAG..............CGCGAAGCG 3450
       ||||||:::   ::::::   :::              ::::::
1085 alLysAspSerLysIleSerAlaAspGlyHisAsnValThrLeuAsnSer 1101

3451 GAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCG 3500
     :::::::::   :::::::::         |||      ::::::   ::
1102 LysValLysThrSerSerSerAsnGlyGlyArgGluSerAsnSerAspAs 1118

3501 GGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACC 3550
     :|||                          :::   :::   ::::::||||:
1118 nAspThrGlyLeuThrIleThrAlaLysAsnValGluValAsnLysAspI 1135

3551 TGATCAGC...CGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACG 3597
     ::   |||   :::   :::|||         ||||||      :::|||
1135 leThrSerLeuLysThrValAsnIleThrAlaSerGluLysValThrThr 1151

3598 CTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGA 3647
      :::::::::   :::         ||||||      |||   :::
1152 .ThrAlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrT 1168

3648 AGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACT 3697
          :::::::         |||      :::::::::|||   ::::::
1168 hrLysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerVal 1184

3698 ACCGTTCGC...AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGC 3744
     ::::::      :::::::::   :::::::::|||      :::
1185 SerAlaThrGlyAspLeuThrThrLysSerGlySerLysIleGluAlaLy 1201

3745 CAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTT 3794
      ::::::   ::::::   |||||||||:::|||:::   ::::::
1201 sSerGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly.... 1216

3795 TTCGCACAACCGGACCGGAAACACCTTCGACGACGGCATCGGCA...... 3838
          |||   :::   :::|||   ::::::||||||:::|||
1217 .....ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1231

3839 ............................ACTCGGCACGGCTTGCCCAC 3858
                                 :::   :::|||      ||
1232 LeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaTh 1248
```

```
     3859 GGTGCCGTTTTCGGGCAAT......ACGGCATCGGCAGGTTCGACATCGG 3902
          |:::      :::||||||      |||::::::||||||||:::   :
     1248 rLeuThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleT 1265

     3903 CATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCA 3952
          ::|||:::::::::::::|||   :::   |||      |||::::::|||
     1265 hrSerThrLysGlyGlnValAspLeuLeuAla......GlnAsnGlySer 1279

     3953 GAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATAC 4002
             |||  |||      |||:::   :::|||             ||
     1280 IleAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyTh 1296

     4003 CGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCG 4052
          |      :::   |||  |||      :::   ||||||:::     :
     1296 rLeuThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuV 1313

     4053 CTATTTCGTCCAAA 4066
          ::|||:::::|||
     1313 alIleAsnAlaLys 1317
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:AAR26503

seq_documentation_block:

```
ID   AAR26503 standard; Protein; 911 AA.
XX
AC   AAR26503;
XX
DT   12-MAR-1993  (first entry)
XX
DE   prn proteins.
XX
KW   B. bronchiseptica; P.68; outer membrane protein; piglet; probe;
KW   atrophic rhinitis; alternative cleavage.
XX
OS   Bordetalla bronchiseptica.
XX
FH   Key             Location/Qualifiers
FT   Protein         35..632
FT                   /label= P.68
FT   Region          266..279
FT                   /label= Repeat_region
FT   Region          570..589
FT                   /label= Repeat_region
FT   Peptide         260..262
FT                   /label= RGD_tripeptide
FT   Peptide         701..703
FT                   /label= RGD_tripeptide
XX
PN   WO9217587-A.
XX
PD   15-OCT-1992.
XX
PF   27-MAR-1992;   92WO-GB00561.
XX
PR   27-MAR-1991;   91GB-0006568.
XX
PA   (WELL ) WELLCOME FOUND LTD.
XX
PI   Charles IG;
XX
DR   WPI; 1992-366258/44.
DR   N-PSDB; AAQ34566.
XX
PT   DNA encoding a Bordetella bronchiseptica protein - used for
PT   obtaining vaccines for preventing respiratory diseases, partic.
PT   atrophic rhinitis in pigs
XX
PS   Claim 1; Fig 1; 28pp; English.
XX
CC   The sequence given is the P.94 antigen from B. bronchiseptica.  The
CC   P.68 antigen is formed by alternative cleavage of this protein.
CC   P.68 is an outer membrane protein with a molecular weight of 68 kD
CC   which is associated with protection of piglets against atrophic
CC   rhinitis.  The DNA sequence encoding these proteins was derived by
CC   standard recombinant DNA techniques using P.68 probes to isolate the
CC   entire P.94 sequence.
XX
SQ   Sequence   911 AA;
```

alignment_scores:

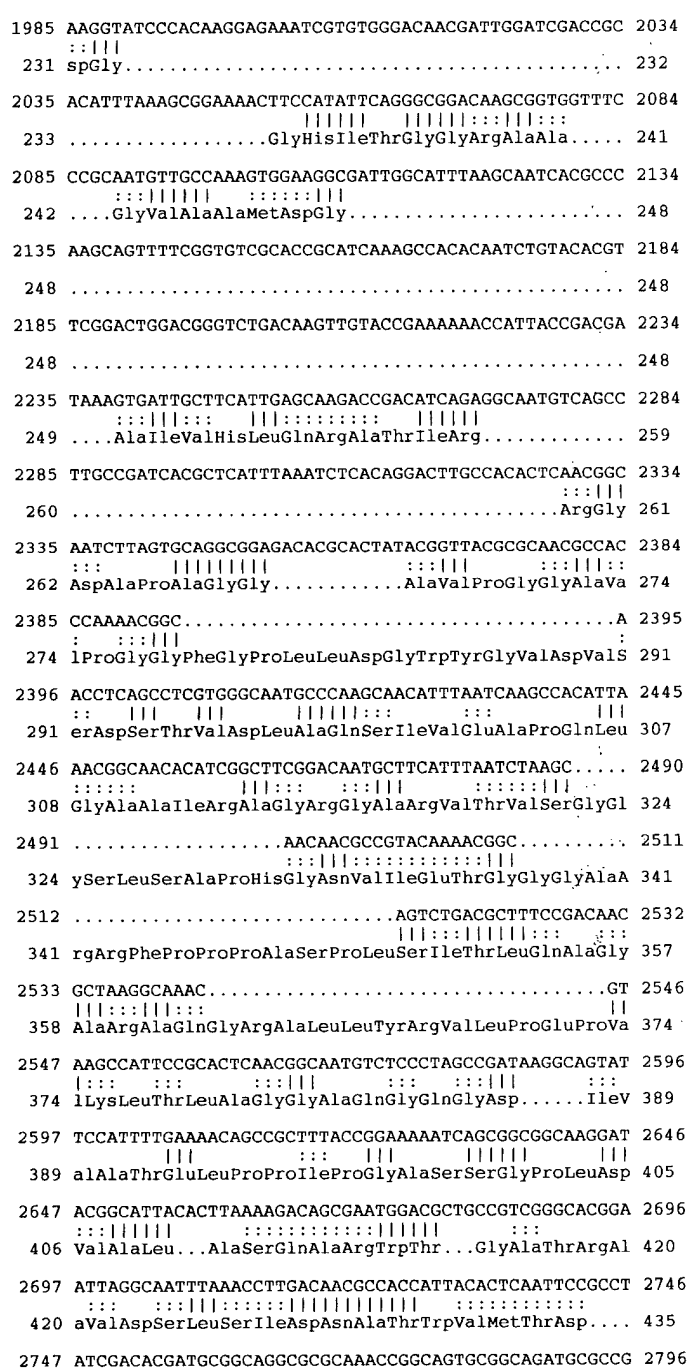
```
            Quality:  279.00              Length:   1149
              Ratio:  0.592                 Gaps:     50
 Percent Similarity:  40.992   Percent Identity: 20.627
```

alignment_block:

US-09-303-518D-653 x AAR26503 ..

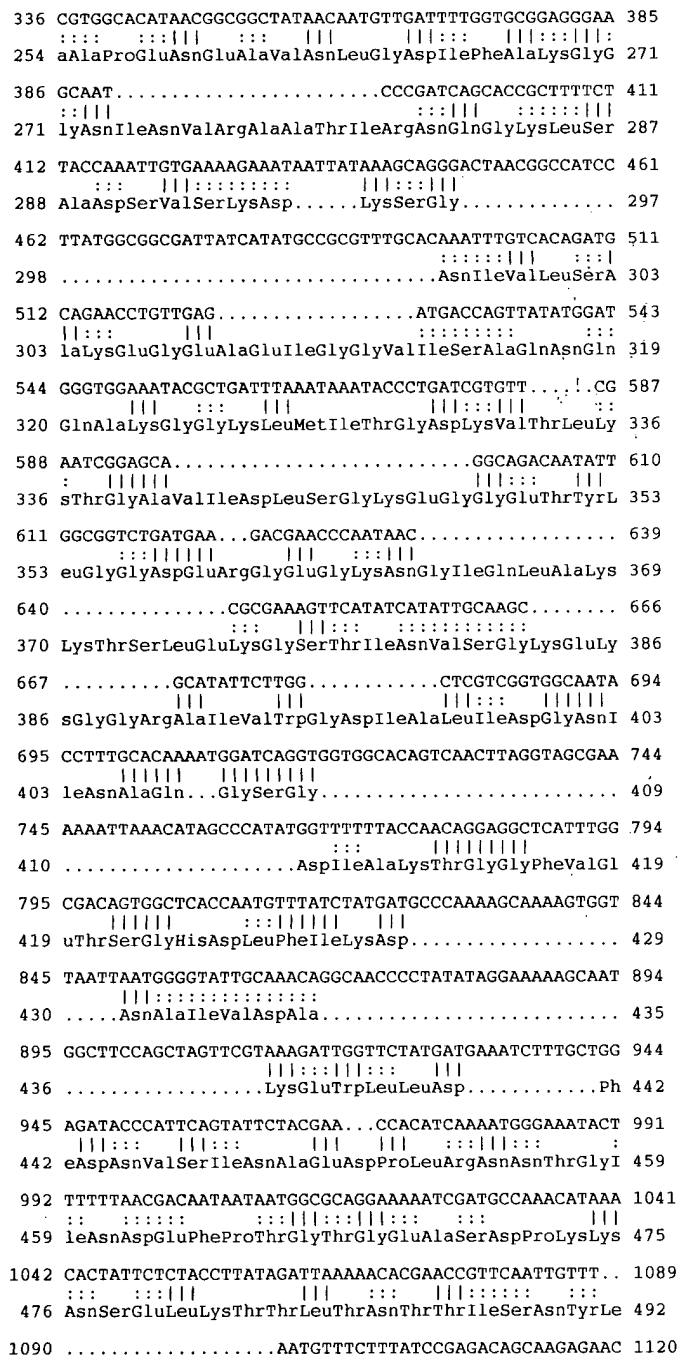
Align seg 1/1 to: AAR26503 from: 1 to: 911

```
    1306 GGCGCGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAA 1355
         |||||||||||||:::   :::   |||:::|||:::      ||||||::
      58 GlyAlaGlyValArgThrAlaThrGlyThrThrIle......LysValSe 72

    1356 CGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGG 1405
         :|||                  :::   ::::::|||:::||||||
      72 rGly.....................ArgGlnAlaGlnGlyValLeuLeuG 82

    1406 TTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAA 1455
           :::      :::|||           :::         :::|||:::
      82 luAsnProAlaAlaGlu............LeuArgPheGlnAsnGlySer 94

    1456 GTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAG 1505
         |||      :::   |||      |||:::|||   ::::::   :::
      95 ValThrSerSerGlyGlnLeuPheAspGluGlyValArgArgPheLeu.. 110

    1506 TGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATA 1555
                                 |||||||||   :::::::|||
     111 .........................GlyThrValThrValLysAla.... 117

    1556 ATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTG 1605
                                                  |||:::|||
     118 ........................................GlyLysLeu 120

    1606 GATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGA 1655
           :::      :::|||:::                 :::   |||::
     121 ValAlaAspHisAlaThrLeuAlaAsnValSerAspThrArgAspAspAs 137

    1656 AGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTT.... 1701
         :|||      :::         :::::::::   ::::::::::::
     137 pGlyIleAlaLeuTyrValAlaGlyGluGlnAlaGlnAlaSerIleAlaA 154

    1702 .....ACCATTACAGGCAATAAAGATATTACTACAACCGGCAATAACAAC 1746
          |||:::   |||         :::   :::      :::   |||
     154 spSerThrLeuGlnGlyAlaGlyGlyValArgValGluArgGlyAlaAsn 170

    1747 AACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAA 1796
               :::::::::   ||||||
     171 ValThrValGlnArgSerThrIleAla..................... 179

    1797 AGATGCAACCAAAACGAACGGGCGGCTCAATCTGAAT.....T 1834
                       :::|||   |||:::::::::
     180 ...............AspGlyGlyLeuHisIleGlyThrLeuGlnProL 191

    1835 ACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAAT 1884
           |||||||||:::         :::   ::::::   |||   |||:::
     191 euGlnProGluAspLeuProProSerArgValValLeuGlyAspThrSer 207

    1885 TTAAACGGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAG 1934
         :::::::::      :::   :::::|||
     208 ValThrAla......ValProAlaSerGly..................... 215

    1935 ACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGG 1984
          :::|||   |||   :::   :::|||::::::         ::::
     216 ....AlaProAlaAlaValSerValPheGlyAlaAsnGluLeuThrValA 231
```


1985 AAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGC 2034
231 spGly 232
2035 ACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTTC 2084
233 GlyHisIleThrGlyGlyArgAlaAla 241
2085 CCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCC 2134
242 GlyValAlaAlaMetAspGly 248
2135 AAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGT 2184
248 248
2185 TCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGACGA 2234
248 248
2235 TAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCC 2284
249 AlaIleValHisLeuGlnArgAlaThrIleArg 259
2285 TTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGC 2334
260 ArgGly 261
2335 AATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCAC 2384
262 AspAlaProAlaGlyGly AlaValProGlyGlyAlaVa 274
2385 CCAAAACGGC A 2395
274 lProGlyGlyPheGlyProLeuLeuAspGlyTrpTyrGlyValAspValS 291
2396 ACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTA 2445
291 erAspSerThrValAspLeuAlaGlnSerIleValGluAlaProGlnLeu 307
2446 AACGGCAACACATCGGCTTCGGACAATGCTTCATTTAATCTAAGC 2490
308 GlyAlaAlaIleArgAlaGlyArgGlyAlaArgValThrValSerGlyGl 324
2491 AACAACGCCGTACAAAACGGC 2511
324 ySerLeuSerAlaProHisGlyAsnValIleGluThrGlyGlyGlyAlaA 341
2512 AGTCTGACGCTTTCCGACAAC 2532
341 rgArgPheProProProAlaSerProLeuSerIleThrLeuGlnAlaGly 357
2533 GCTAAGGCAAAC GT 2546
358 AlaArgAlaGlnGlyArgAlaLeuLeuTyrArgValLeuProGluProVa 374
2547 AAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTAT 2596
374 lLysLeuThrLeuAlaGlyGlyAlaGlnGlyGlnGlyAsp IleV 389
2597 TCCATTTTGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGAT 2646
389 alAlaThrGluLeuProProIleProGlyAlaSerSerGlyProLeuAsp 405
2647 ACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGA 2696
406 ValAlaLeu AlaSerGlnAlaArgTrpThr GlyAlaThrArgAl 420
2697 ATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCT 2746
420 aValAspSerLeuSerIleAspAsnAlaThrTrpValMetThrAsp 435
2747 ATCGACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCG 2796
436 AsnSerAsnValGlyAlaLeu 442
2797 CGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGC 2846
443 ArgLeuAlaSerAspGlySer ValAspPheGlnGlnProAlaGluAl 458
2847 AGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGG 2896
458 a GlyArgPheLysCysLeuMetValAsp ThrLeuAlaGlySerG 473
2897 GAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTG 2946
473 lyLeuPheArgMetAsnValPheAlaAspLeuGlyLeuSerAspLysLeu 489
2947 AAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTCAACAATAC 2996
490 ValValMetArgAspAlaSerGlyGlnHisArgLeuLeuValArgAsnSe 506
2997 CGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAG 3046
506 rGlySerGluProAlaSerGlyAsnThrMetLeuLeuValGln 520
3047 ACAACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAA 3093
521 ThrProArgGlySerAlaAlaThrPheThrLeuAlaAsnLysAsp 535
3094 CACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAA GA 3137
536 GlyLysValAspIleGlyThrTyrArgTyrArgLeuAlaAlaAsnGlyAs 552
3138 CGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACA 3187
552 nGlyGlnTrpSerLeu.Val 558
3188 AACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCA 3237
558 558
3238 CAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGA 3287
558 558
3288 CGCGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTG 3337
559 GlyAlaLysAlaProProAlaProLysProAla 569
3338 CCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGCCGGCATTATGCAGGCG 3387
570 Pro GlnProGlyProGln 575
3388 GAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAA 3437
575 575
3438 ACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCC 3487
576 ProGlyProGlnProProGln ProProGlnPr 586
3488 GCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAA 3537
586 oProGlnProProGlnArgGlnProGluAla ProAlaProGlnProPro 602
3538 CCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATT 3587
603 AlaGlyArgGluLeu SerAlaAlaAlaAsnAlaAlaValAsnThrGl 618
3588 TTCCGCCACGCTCAACAGCGTT TTCGCCGTACAGGACGAATTGG 3631
618 yGlyValGlyLeuAlaSerThrLeuTrpTyrAlaGluSerAsnAlaLeuS 635
3632 ACCGCGTGTTTGCCGAAGACCGCCGCAAC GCCGTTTGG 3669

```
 635 erLysArgLeuGlyGluLeuArgLeuAsnProAspAlaGlyGlyAlaTrp 651

3670 ACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGC 3719
          |||:::                                    ||
 652 GlyArgGlyPhe.....................................Al 656

3720 CTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCG 3769
     |   |||||||||   |||
 656 aGlnArgGlnGlnLeuAsp.............................. 662

3770 GCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACACC 3819
                                    ||||||:::|||:::
 663 ................................AsnArgAlaGlyArgArg 668

3820 TTCGACGAC.....................GGCATCGGCAACTCGGCACG 3848
     |||||||:::                    |||      :::::::::
 669 PheAspGlnLysValAlaGlyPheGluLeuGlyAlaAspHisAlaValAl 685

3849 GCTTGCCCACGGTGCC......GTTTTCGGGCAATACGGCATCGGCAGGT 3892
      :::|||   |||         :::   |||      |||      |||
 685 aValAlaGlyGlyArgTrpHisLeuGlyGlyLeuAlaGlyTyrThrArgG 702

3893 TCGACATCGGCATCAGC.................................. 3909
       |||   |||::::::
 702 lyAspArgGlyPheThrGlyAspGlyGlyGlyHisThrAspAsnValAsn 718

3910 .......................GCGGGCGCGGGTTTT............ 3924
                             |||:::::||||||
 719 ValGlyGlyTyrProThrTyrIleAlaAsnSerGlyPheTyrLeuAspAl 735

3925 .....................................AGTAGCGGCAGCC 3937
                                         :::||||||
 735 aThrLeuArgAlaSerArgLeuGluAsnAspPheLysValAlaGlySerA 752

3938 TTTCAGACGGCATCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGC 3987
       :::   :::::::::||||||   |||          ~   :::|||
 752 spGlyTyrAlaValLysGlyLysTyrArgThr............HisGly 764

3988 ATTCAGGCAAGATACCGCGCAGGT..............TTCGGCGGATT 4022
     :::   |||      :::|||||||                  |||::
 765 ValGlyAlaSerLeuGluAlaGlyArgArgPheAlaHisAlaAspGlyTr 781

4023 CGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGG 4072
     :   :::|||||||:::         :::   :::   |||   :::::::
 781 pPheLeuGluProGlnAlaGluLeuAlaValPheArgValGlyArgGlyS 798

4073 ATTACCGATACGAAAAC...GTCAATATCGCCACCCCGGGCCTTGCATTC 4119
     ::|||+|||      |||   :::::::::        |||   :::
 798 erTyrArgAlaAlaAsnGlyLeuArgValArgAspGluGlyGlySerSer 814

4120 AACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCA 4169
           |||   |||:::::::::        ::::::   |||
 815 ValLeuGlyArgLeuGlyLeuGluValGlyLysArgIleGluLeuAlaGl 831

4170 ACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTT 4219
           ::::::   ||||||::::::   |||      :::      :
 831 yGlyArgGlnValGlnProTyrIleLysAlaSerValLeuGlnGluPheA 848

4220 CCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCGCAGGATTTC 4269
     ::|||         :::|||:::|||     :::       ::::::
 848 spGlyAlaGlyThrValArgThrAsnGlyIleAlaHisArgThrGluLeu 864

4270 GGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTT 4319
          ||||||   ||||||   |||
 865 ArgGlyThrArg...AlaGluLeuGly..................... 872

4320 CACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCG. 4368
         |||::::::   ||||||   :::::::|||   :::|||   |||
 873 ....LeuGlyMetAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrAlaS 888

4369 .....CAGCACAGCGCGGGCATCAAATTAGGCTACCGCTGG 4404
          ::::::|||   |||   ||||||:::      |||
 888 erTyrGluTyrSerLysGlyProLysLeuAlaMetProTrp 901
```

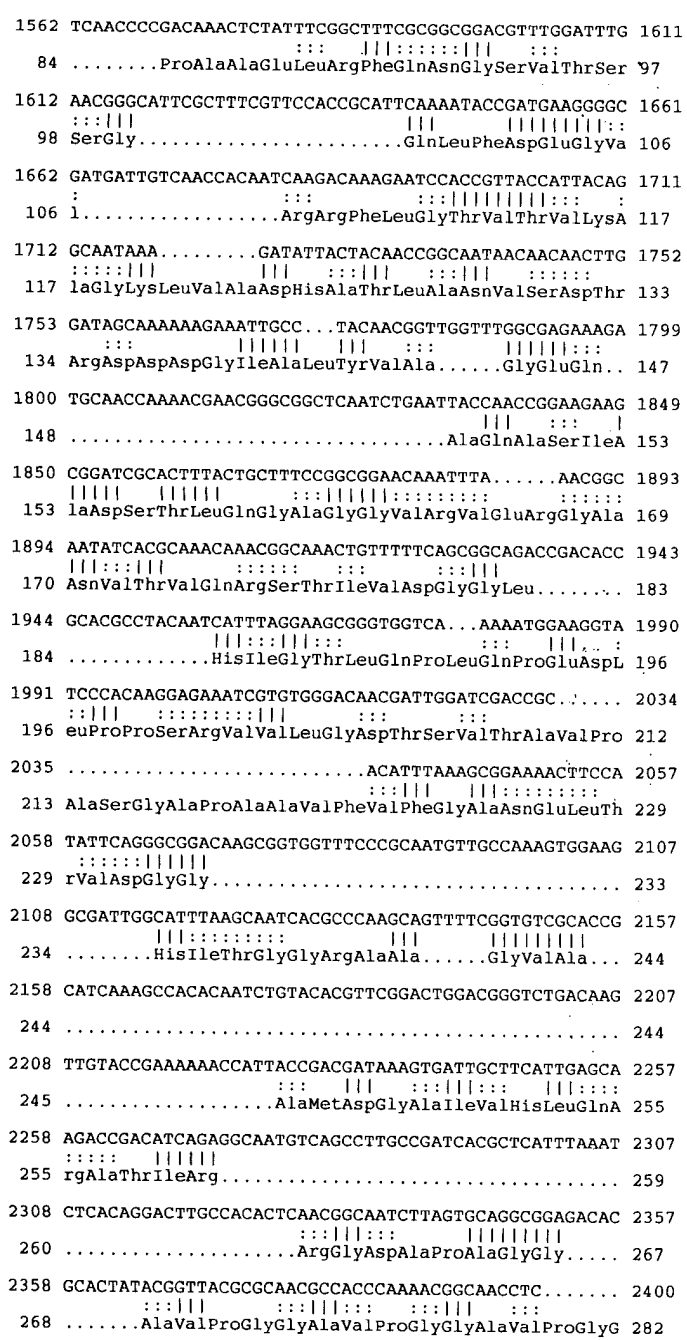
```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41724

seq_documentation_block:
ID   AAR41724 standard; Protein; 1477 AA.
XX
AC   AAR41724;
XX
DT   26-APR-1994  (first entry)
XX
DE   High molecular weight protein 2 (HMW2).
XX
KW   HMW; high molecular weight protein; virus; vaccine; influenza;
KW   epitope; immunity; haemophilus influenzae.
XX
OS   Haemophilus influenzae.
XX
PN   WO9319090-A.
XX
PD   30-SEP-1993.
XX
PF   16-MAR-1993;   93WO-US02166.
XX
PR   16-MAR-1992;   92GB-0005704.
XX
PA   (BARE/) BARENKAMP S J.
PA   (INRM ) INSERM INST NAT SANTE & RECH  MEDICALE.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1993-320683/40.
DR   N-PSDB; AAQ49507.
XX
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties
XX
PS   Claim 4; Figure 4; 100pp; English.
XX
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.
XX
SQ   Sequence   1477 AA;


alignment_scores:
           Quality:  278.50              Length:   1488
             Ratio:  0.425                 Gaps:     73
 Percent Similarity:  44.086   Percent Identity: 19.220

alignment_block:
 US-09-303-518D-653 x AAR41724   ..

 Align seg 1/1  to: AAR41724  from: 1  to: 1477

     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
         |||   |||   ::::::||||||   ||||||   :::|||
     199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal........ 211

     228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
                ::::|||:::|||   ::::::   :::
     212 .......AsnLeuIleGlyGlyLysValLysAsnGlu............. 221

     278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
           :::|||:::|||   ||||||      ::::::   |||||||||:::
     222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

     322 CAA.................................TATATTGTGAG 335
         :::                                 |||   :::::
     238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254
```


336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271
386 GCAAT.....CCCGATCAGCACCGCTTTTCT 411
271 lyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287
412 TACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCC 461
288 AlaAspSerValSerLysAsp......LysSerGly..... 297
462 TTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATG 511
298 .....AsnIleValLeuSerA 303
512 CAGAACCTGTTGAG.....ATGACCAGTTATATGGAT 543
303 laLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319
544 GGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT....!.CG 587
320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLy 336
588 AATCGGAGCA.....GGCAGACAATATT 610
336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353
611 GGCGGTCTGATGAA...GACGAACCCAATAAC..... 639
353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369
640 .....CGCGAAAGTTCATATCATATTGCAAGC..... 666
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLy 386
667 .....GCATATTCTTGG.....CTCGTCGGTGGCAATA 694
386 sGlyGlyArgAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnI 403
695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAA 744
403 leAsnAlaGln...GlySerGly..... 409
745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
410 .....AspIleAlaLysThrGlyGlyPheValGl 419
795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGT 844
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAAT 894
430 .....AsnAlaIleValAspAla..... 435
895 GGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGG 944
436 .....LysGluTrpLeuLeuAsp.....Ph 442
945 AGATACCCATTCAGTATTCTACGAA...CCACATCAAAATGGGAAATACT 991
442 eAspAsnValSerIleAsnAlaGluAspProLeuArgAsnAsnThrGlyI 459
992 TTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAA 1041
459 leAsnAspGluPheProThrGlyThrGlyGluAlaSerAspProLysLys 475
1042 CACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTT.. 1089
476 AsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnTyrLe 492
1090 .....AATGTTTCTTTATCCGAGACAGCAAGAGAAC 1120
492 uLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr.... 507
1121 CTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAAT 1170
508 .....ValAsnSer.....SerIleAsn 513
1171 AATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACT 1220
514 IleGlySerAsnSerHisLeuIleLeuHisSerLysGly..... 526
1221 TACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATT 1270
527 .....GlnArgGlyGlyGlyValGlnIleAspGlyAspI 538
1271 TTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 1320
538 leThrSerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrpValAsp 554
1321 ATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGA 1370
555 ValHisLysAsnIleThr..... 560
1371 CCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCC..... 1413
561 .....LeuAspGlnGlyPheLeuAsnIleThrAlaAlaSerV 573
1414 .....AAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAA 1455
573 alAlaPheGluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLys 589
1456 GTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAG 1505
590 IleVal..... 591
1506 TGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATA 1555
592 .....AlaGlnGlyThrValThrIleThrGlyGluG 602
1556 ATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTG 1605
602 lyLys.....AspPheArgAlaAsnAsnVal 610
1606 GATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGA 1655
611 SerLeuAsnGly.....ThrGlyLy 617
1656 AGGGGCGATGATTGTC.....AACCACAATCAAGACAAAGAAT 1693
617 sGlyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerG 634
1694 CCACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGCAATAAC 1743
634 lyThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThrThrArgLys 650
1744 AAC..... 1746
651 AsnThrSerTyrTrpGlnThrSerHisAspSerHisTrpAsnValSerAl 667
1747 ....AACTTGGAT..... 1755
667 aLeuAsnLeuGluThrGlyAlaAsnPheThrPheIleLysTyrIleSerS 684
1756 .....AGCAAAAAAGAAATTGCC 1773
684 erAsnSerLysGlyLeuThrThrGlnTyrArgSerSerAlaGlyValAsn 700
1774 TACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCT 1823
701 PheAsnGly.....ValAsnGlyAsnMe 708
1824 CAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCG 1873

```
    3503 ATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTG 3552
                                              ::::::|||::::
    1132 ..................................AsnLysAspIle 1135

    3553 ATCAGC...CGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCT 3599
           |||   :::   :::|||         ||||||      :::||| :
    1136 ThrSerLeuLysThrValAsnIleThrAlaSerGluLysValThrThr.T 1152

    3600 CAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAG 3649
         ::::::::   :::         ||||||      |||   :::
    1152 hrAlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThr 1168

    3650 ACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTAC 3699
         ::::::         |||      :::::::::|||   ::::::::
    1169 LysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSe 1185

    3700 CGTTCGC...AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCA 3746
         ::::         :::::::::   :::::::::|||      :::   :
    1185 rAlaThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysS 1202

    3747 AATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTT 3796
         :::::   ::::::   |||||||||:::|||:::   ::::::
    1202 erGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly...... 1216

    3797 CGCACAACCGGACCGGAAACACCTTCGACGACGGCATCGGCA........ 3838
           |||   :::   :::|||   ::::::||||||:::|||
    1217 ...ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLe 1232

    3839 ..........................ACTCGGCACGGCTTGCCCACGG 3860
                                     :::   :::|||      |||:
    1232 uThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrL 1249

    3861 TGCCGTTTTCGGGCAAT......ACGGCATCGGCAGGTTCGACATCGGCA 3904
         ::      :::||||||      |||:::::|||||||||:::   :::
    1249 euThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThr 1265

    3905 TCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGA 3954
         |||:::::::::::::|||   :::   |||      |||::::::|||
    1266 SerThrLysGlyGlnValAspLeuLeuAla......GlnAsnGlySerIl 1280

    3955 GGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCG 4004
          |||   |||      |||:::   :::|||               |||
    1280 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrL 1297

    4005 CGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCT 4054
           :::   |||   |||      :::   ||||||:::      :::
    1297 euThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuVal 1313

    4055 ATTTCGTCCAAA 4066
         |||::::::|||
    1314 IleAsnAlaLys 1317
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:AAR25578

seq_documentation_block:
ID   AAR25578 standard; Protein; 922 AA.
XX
AC   AAR25578;
XX
DT   08-JAN-1993  (first entry)
XX
DE   Bordetella parapertussis P95 antigen precursor.
XX
KW   Whooping cough; P70 antigen; P95 precursor protein; vaccination.
XX
OS   Bordetella parapertussis.
XX
FH   Key             Location/Qualifiers
FT   Protein         35..643
FT                   /label= P70
FT   Binding-site    260..262
FT                   /note= "motif associated with cell-cell adhesion"
FT   Region          266..285
FT                   /note= "contains 5 direct, tandem repeats"
FT   Region          575..612
FT                   /note= "contains 9 direct repeats of Pro-Gln-Pro"
FT   Binding-site    712..714
FT                   /note= "motif associated with cell-cell adhesion"
XX
PN   WO9211292-A.
XX
PD   09-JUL-1992.
XX
PF   23-DEC-1991;   91WO-GB02302.
XX
PR   21-DEC-1990;   90GB-0027901.
XX
PA   (WELL ) WELLCOME FOUND LTD.
XX
PI   Charles IG;
XX
DR   WPI; 1992-250033/30.
DR   N-PSDB; AAQ26509.
XX
PT   Acellular vaccine for immunisation against whooping cough -
PT   comprises protein uncontaminated by B. para:pertussis components
PT   and capable of binding antibodies which bind native P70 antigen
XX
PS   Claim 1; Fig 1; 20pp; English.
XX
CC   A cosmid library was constructed by transforming E.coli HB101 with
CC   recombinant cosmids prepared by partial digestion of B.parapertussis
CC   chromosomal DNA with Sau3A and cloning of 40-50kb fragments into the
CC   BamHI site of cosmid pHC79. The cosmids were screened with a 1.8kb
CC   ClaI fragment from the prn gene of B.pertussis. The insert from one
CC   positive colony, harbouring cosmid pBD811, was sequenced and found to
CC   contain an open reading frame encoding a 922 amino acid protein
CC   with calculated mol.wt. 95,177. This precursor protein ("P95") is
CC   processed in vivo to the P70 antigen of apparent mol. wt. 70,000 as
CC   determined by SDS-PAGE, but with actual mol.wt. 61kD. Antigenic
CC   fragments of the protein will be useful in developing an acellular
CC   vaccine against B.parapertussis. Preferred fragments include amino
CC   acids Pro577 to Pro612 or Ala574 to Pro612.
XX
SQ   Sequence   922 AA;
```

```
alignment_scores:
           Quality:  275.50            Length:   1162
             Ratio:  0.566               Gaps:     54
 Percent Similarity:  41.910   Percent Identity: 21.687

alignment_block:
 US-09-303-518D-653 x AAR25578   ..

 Align seg 1/1  to: AAR25578  from: 1  to: 922

    1366 AACGACCGCCTGTCCAAAATCGGCAAAGGC.............ACGCT 1400
         |||      ||||||:::|||   |||:::              |||
       2 AsnMetSerLeuSerArgIleValLysAlaAlaProLeuArgArgThrTh 18

    1401 GCTGGTTCAAGCCAAAGGG.............................. 1419
          |||::::::|||   |||
      18 rLeuAlaMetAlaLeuGlyAlaLeuGlyAlaAlaProAlaAlaTyrAlaA 35

    1420 .....GAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAA...GTCATC 1461
              :::||||||:::   :::::::::|||:::   :::      |||
      35 spTrpAsnAsnGlnSerIleIleLysAlaGlyGluArgGlnHisGlyIle 51

    1462 TTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAAT 1511
               :::|||:::|||      |||   :::   |||   :::
      52 HisIleLysGlnSerAspGlyAlaGlyValArgThrAlaThrGlyThrTh 68

    1512 CGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAGT 1561
               ||||||||||||   :::      ::      :::|||
      68 rIleLysValSerGlyArgGlnAlaGlnGlyValLeuLeuGluAsn.... 83
```


1562 TCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTG 1611
84 .....ProAlaAlaGluLeuArgPheGlnAsnGlySerValThrSer 97
1612 AACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGC 1661
98 SerGly.....GlnLeuPheAspGluGlyVa 106
1662 GATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAG 1711
106 l.....ArgArgPheLeuGlyThrValThrValLysA 117
1712 GCAATAAA.....GATATTACTACAACCGGCAATAACAACAACTTG 1752
117 laGlyLysLeuValAlaAspHisAlaThrLeuAlaAsnValSerAspThr 133
1753 GATAGCAAAAAAGAAATTGCC...TACAACGGTTGGTTTGGCGAGAAAGA 1799
134 ArgAspAspAspGlyIleAlaLeuTyrValAla.....GlyGluGln.. 147
1800 TGCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAG 1849
148 .....AlaGlnAlaSerIleA 153
1850 CGGATCGCACTTTACTGCTTTCCGGCGGAACAAATTTA.....AACGGC 1893
153 laAspSerThrLeuGlnGlyAlaGlyGlyValArgValGluArgGlyAla 169
1894 AATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACC 1943
170 AsnValThrValGlnArgSerThrIleValAspGlyGlyLeu..... 183
1944 GCACGCCTACAATCATTTAGGAAGCGGGTGGTCA...AAAATGGAAGGTA 1990
184 .....HisIleGlyThrLeuGlnProLeuGlnProGluAspL 196
1991 TCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGC..... 2034
196 euProProSerArgValValLeuGlyAspThrSerValThrAlaValPro 212
2035 .....ACATTTAAAGCGGAAAACTTCCA 2057
213 AlaSerGlyAlaProAlaAlaValPheValPheGlyAlaAsnGluLeuTh 229
2058 TATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAG 2107
229 rValAspGlyGly..... 233
2108 GCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGCACCG 2157
234 .....HisIleThrGlyGlyArgAlaAla.....GlyValAla... 244
2158 CATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAG 2207
244 244
2208 TTGTACCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGAGCA 2257
245 .....AlaMetAspGlyAlaIleValHisLeuGlnA 255
2258 AGACCGACATCAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAAT 2307
255 rgAlaThrIleArg..... 259
2308 CTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGCAGGCGGAGACAC 2357
260 .....ArgGlyAspAlaProAlaGlyGly..... 267
2358 GCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACCTC..... 2400
268 .....AlaValProGlyGlyAlaValProGlyGlyAlaValProGlyG 282
2401 .....AGC 2403
282 lyPheGlyProLeuLeuAspGlyTrpTyrGlyValAspValSerAspSer 298
2404 CTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAA 2453
299 ThrValAspLeuAlaGlnSerIleValGluAlaProGlnLeuGlyAlaAl 315
2454 CACATCGGCTTCGGACAATGCTTCATTTAATCTAAGC..... 2490
315 aIleArgAlaGlyArgGlyAlaArgValThrValSerGlyGlySerLeuS 332
2491 .....AACAACGCCGTACAAAACGGC..... 2511
332 erAlaProHisGlyAsnValIleGluThrGlyGlyGlyAlaArgArgPhe 348
2512 .....AGTCTGACGCTTTCCGACAACGCTAAGGC 2540
349 ProProProAlaSerProLeuSerIleThrLeuGlnAlaGlyAlaArgAl 365
2541 AAAC.....GTAAGCCATT 2554
365 aGlnGlyArgAlaLeuLeuTyrArgValLeuProGluProValLysLeuT 382
2555 CCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTT 2604
382 hrLeuAlaGlyGlyAlaGlnGlyGlnGlyAsp.....IleValAlaThr 396
2605 GAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATT 2654
397 GluLeuProProIleProGlyAlaSerSerGlyProLeuAspValAlaLe 413
2655 ACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCA 2704
413 u...AlaSerGlnAlaArgTrpThr...GlyAlaThrArgAlaValAspS 428
2705 ATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACAC 2754
428 erLeuSerIleAspAsnAlaThrTrpValMetThrAsp..... 440
2755 GATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCG 2804
441 .....AsnSerAsnValGlyAlaLeuArgLeuAl 450
2805 TTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCC 2854
450 aSerAspGlySer...ValAspPheGlnGlnProAlaGluAla...GlyA 465
2855 GTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTC 2904
465 rgPheLysValLeuMetValAsp...ThrLeuAlaGlySerGlyLeuPhe 480
2905 CGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGC 2954
481 ArgMetAsnValPheAlaAspLeuGlyLeuSerAspLysLeuValValMe 497
2955 GGAAAGTTCCGAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAACG 3004
497 tArgAspAlaSerGlyGlnHisArgLeuTrpValArgAsnSerGlySerG 514
3005 AACCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACA 3054
514 luProAlaSerGlyAsnThrMetLeuLeuValGln.....Thr 526
3055 CCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAA.....CACGT 3098
527 ProArgGlySerAlaAlaThrPheThrLeuAlaAsnLysAspGlyLysVa 543
3099 CGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAA...GACGGCGAGT 3145
543 lAspIleGlyThrTyrArgTyrArgLeuAlaAlaAsnGlyAsnGlyGlnT 560
3146 TCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGC 3195

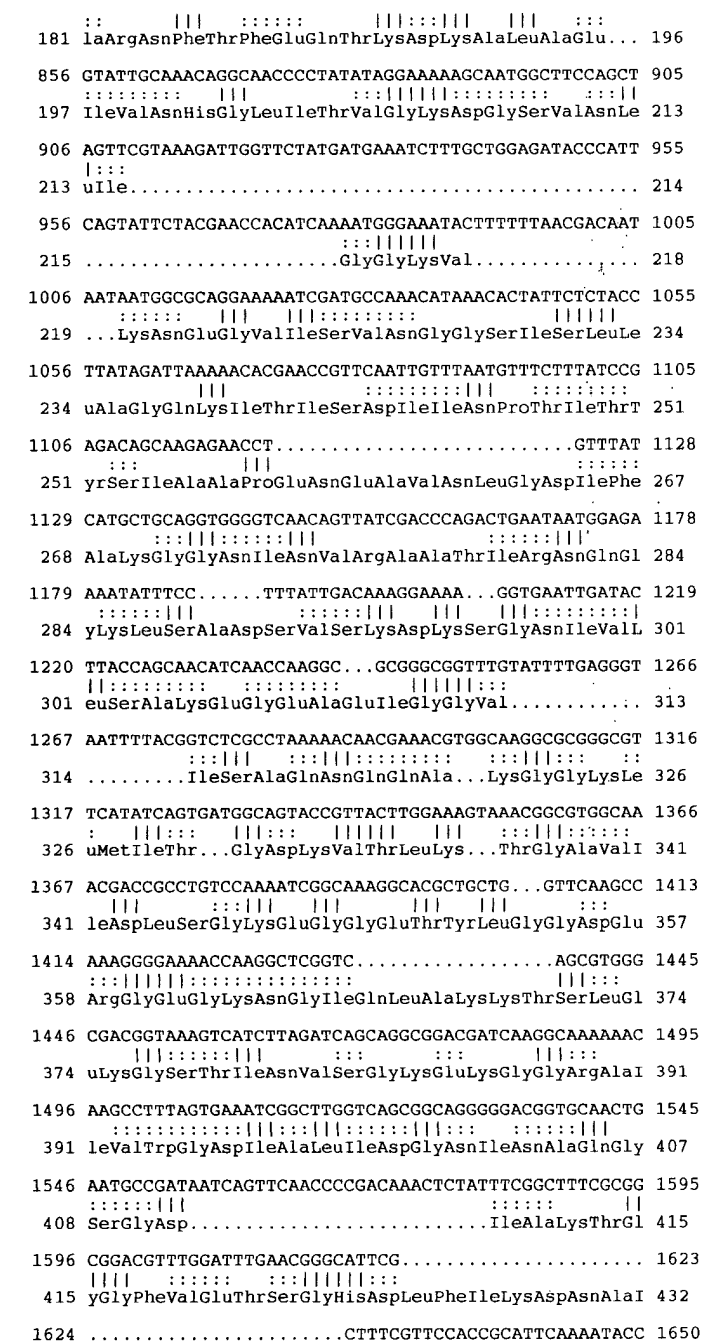
```
      ::   ||| :::
 560 rpSerLeu.Val..................................... 563

3196 AAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGC 3245

 563 .................................................. 563

3246 CGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGA 3295

 563 .................................................. 563

3296 TTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCG 3345
              ||||||    |||          |||      |||   ::
 564 .........GlyAlaLysAlaProProAlaProLysProAlaPro...Gl 576

3346 GCCCGGCAGGCAGGCGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGA 3395
      :||||||   |||   |||   :::|||
 576 nProGlyProGlnProGlyProGlnProPro................... 586

3396 GAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCG 3445

 586 .................................................. 586

3446 AAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCC 3495
                       ||||||    |||||:::||||||:::|
 587 ..................GlnProProGln.ProProGlnProProGlnP 597

3496 CGCCGGGATTTGCCGCAACCGCAGCCCCAA.........CCGCAACCCCA 3536
     ||||| :::    ||||||   ||||||:::        |||||||||
 597 roPro.GlnProProGlnArgGlnProGluAlaProAlaProGlnProPr 613

3537 ACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAAT 3586
             |||:::|||   |||      ||||||:::::::::::
 613 oAlaGlyArgGluLeu...SerAlaAlaAlaAsnAlaAlaValAsnThrG 629

3587 TTTCCGCCACGCTCAACAGCGTT......TTCGCCGTACAGGACGAATTG 3630
       ::::::   |||   |||:::      :::|||   ::::::   |||
 629 lyGlyValGlyLeuAlaSerThrLeuTrpTyrAlaGluSerAsnAlaLeu 645

3631 GACCGCGTGTTTGCCGAAGACCGCCGCAAC...........GCCGTTTG 3668
     ::::::   ::::::|||   |||   |||            ::::::||
 646 SerLysArgLeuGlyGluLeuArgLeuAsnProAspAlaGlyGlyAlaTr 662

3669 GACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCG 3718
     |       |||:::                                  |
 662 pGlyArgGlyPhe....................................A 667

3719 CCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTC 3768
      ||    |||||||||   |||
 667 laGlnArgGlnGlnLeuAsp.............................. 673

3769 GGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACAC 3818
                                       ||||||:::|||:::
 674 ..................................AsnArgAlaGlyArgAr 679

3819 CTTCGACGAC....................GGCATCGGCAACTCGGCAC 3847
      ||||||:::                    |||   :::::::::
 679 gPheAspGlnLysValAlaGlyPheGluLeuGlyAlaAspHisAlaValA 696

3848 GGCTTGCCCACGGTGCC......GTTTTCGGGCAATACGGCATCGGCAGG 3891
       :::|||   |||          :::   |||      |||      |||
 696 laValAlaGlyGlyArgTrpHisLeuGlyGlyLeuAlaGlyTyrThrArg 712

3892 TTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTT.. 3939
        |||   |||:::::::::   :::|||      :::   |||:::
 713 GlyAspArgGlyPheThrGlyAspGlyGlyGlyHisThrAspSerValHi 729

3939 .................................................. 3939

 729 sValGlyGlyTyrAlaThrTyrIleAlaAsnSerGlyPheTyrLeuAspA 746

3940 ...............................................TCA 3942
                                                    |||
 746 laThrLeuArgAlaSerArgLeuGluAsnAspPheLysValAlaGlySer 762

3943 GACGGC......ATCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGG 3986
     ||||||      ::::::||||||   |||   :::   :::      ::
 763 AspGlyTyrAlaValLysGlyLysTyrArgThrHisGlyValGlyValSe 779

3987 CATTCAGGCAAGATACCGC...GCAGGTTTCGGCGGATTCGGCATCGAAC 4033
     ::::::|||      |||   |||         |||:::   :::||||
 779 rLeuGluAlaGlyArgArgPheAlaHisAlaAspGlyTrpPheLeuGluP 796

4034 CGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGATAC 4083
     ||:::         :::   :::   |||       :::   ||||||
 796 roGlnAlaGluLeuAlaValPheArgValGlyGlyGlyAlaTyrArgAla 812

4084 GAAAAC...GTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCTACCG 4130
       |||   :::::::::            |||   :::           ||
 813 AlaAsnGlyLeuArgValArgAspGluGlyGlySerSerValLeuGlyAr 829

4131 CGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCA 4180
     |   |||:::::::::             ::::::   |||        :::
 829 gLeuGlyLeuGluValGlyLysArgIleGluLeuAlaGlyGlyArgGlnV 846

4181 TCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCCGGCAAAGTC 4230
     ::   ||||||::::::   |||         :::        :::|||
 846 alGlnProTyrIleLysAlaSerValLeuGlnGluPheAspGlyAlaGly 862

4231 CGAACGCGCGTCAATACCGCCGTATTGGCGCAGGATTTCGGCAAAACCCG 4280
       :::|||:::|||            :::        ::::::      |||||
 863 ThrValArgThrAsnGlyIleAlaHisArgThrGluLeuArgGlyThrAr 879

4281 CAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCC 4330
     |   ||||||   |||                            |||:::
 879 g...AlaGluLeuGly..........................LeuGlyM 886

4331 TCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCG......CAGCAC 4374
     ::   ||||||   ::::::|||   :::|||   |||      ::::::
 886 etAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrAlaSerTyrGluTyr 902

4375 AGCGCGGGCATCAAATTAGGCTACCGCTGG 4404
     |||   |||   ||||||:::      |||
 903 SerLysGlyProLysLeuAlaMetProTrp 912
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:AAR14320

```
seq_documentation_block:
ID   AAR14320 standard; Protein; 911 AA.
XX
AC   AAR14320;
XX
DT   20-JAN-1992  (first entry)
XX
DE   Pertactin antigen P.68.
XX
KW   Pertactin; Pichia; B. pertussis; B. parapertussis.
XX
OS   Bordetella bronchiseptica.
XX
FH   Key             Location/Qualifiers
FT   Peptide         266..270
FT                   /label= repeat
FT   Peptide         271..275
FT                   /label= repeat
FT   Peptide         570..572
FT                   /label= repeat
FT   Peptide         574..576
FT                   /label= repeat
FT   Peptide         578..580
FT                   /label= repeat
FT   Peptide         581..583
FT                   /label= repeat
FT   Peptide         584..586
FT                   /label= repeat
FT   Peptide         587..589
```

```
FT                   /label= repeat
FT   Peptide         599..601
FT                   /label= repeat
XX
PN   WO9115571-A.
XX
PD   17-OCT-1991.
XX
PF   28-MAR-1991;   91WO-GB00487.
XX
PR   02-APR-1990;   90GB-0007416.
XX
PA   (WELL ) WELLCOME FOUNDATION LTD.
XX
PI   Clare JJ,  Romanos MA;
XX
DR   WPI; 1991-325214/44.
DR   N-PSDB; AAQ14319.
XX
PT   Pichia microorganism transformants - for production of
PT   Bordetella pertactin antigens for whooping cough vaccines
XX
PS   Disclosure; Fig 1B; 38pp; English.
XX
CC   Pichia microorganisms are transformed for the expression of
CC   pertactin antigens.  DNA sequence used are represented in AAQ14319-20
CC   encoding the B. bronchiseptica P.68 and B. parapertussis P.70 antigen
CC   respectively or the B. pertussis P.69 encoding sequence described
CC   by I.G. Charles et al. Proc. Natl. Acad. Sci. USA, Vol. 80:3554-3448
CC   (1989).
XX
SQ   Sequence   911 AA;


alignment_scores:
            Quality:  274.00              Length:   1149
              Ratio:  0.584                Gaps:     50
 Percent Similarity:  40.818   Percent Identity: 20.540

alignment_block:
 US-09-303-518D-653 x AAR14320   ..

 Align seg 1/1  to: AAR14320  from: 1  to: 911

    1306 GGCGCGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAA 1355
         ||||||||||||:::   :::   |||:::|||:::      ||||||::
      58 GlyAlaGlyValArgThrAlaThrGlyThrThrIle......LysValSe 72

    1356 CGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGG 1405
         :|||                  :::   ::::::|||:::||||||
      72 rGly.....................ArgGlnAlaGlnGlyValLeuLeuG 82

    1406 TTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAA 1455
           :::      :::|||           :::         :::|||:::
      82 luAsnProAlaAlaGlu............LeuArgPheGlnAsnGlySer 94

    1456 GTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAG 1505
         |||      :::   |||      |||:::|||   ::::::   :::
      95 ValThrSerSerGlyGlnLeuPheAspGluGlyValArgArgPheLeu.. 110

    1506 TGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATA 1555
                              |||||||||   ::::::|||
     111 .........................GlyThrValThrValLysAla.... 117

    1556 ATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTG 1605
                                          |||:::|||
     118 ........................................GlyLysLeu 120

    1606 GATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGA 1655
              :::      :::|||:::              :::   |||::
     121 ValAlaAspHisAlaThrLeuAlaAsnValSerAspThrArgAspAspAs 137

    1656 AGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTT.... 1701

         :|||      :::         :::::::::   ::::::::::::
     137 pGlyIleAlaLeuTyrValAlaGlyGluGlnAlaGlnAlaSerIleAlaA 154

    1702 .....ACCATTACAGGCAATAAAGATATTACTACAACCGGCAATAACAAC 1746
               |||:::   |||          :::   :::      :::   |||
     154 spSerThrLeuGlnGlyAlaGlyGlyValArgValGluArgGlyAlaAsn 170

    1747 AACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAA 1796
                  ::::::
     171 ValThrValGlnArg..................................... 175

    1797 AGATGCAACCAAAACGAACGGGCGGCTCAATCTGAAT.............T 1834
              :::|||   ::::::::|||   |||::::::::::
     176 ....SerThrIleValAspGlyGlyLeuHisIleGlyThrLeuGlnProL 191

    1835 ACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAAT 1884
           |||||||||:::         :::   ::::::   |||   |||:::
     191 euGlnProGluAspLeuProProSerArgValValLeuGlyAspThrSer 207

    1885 TTAAACGGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAG 1934
         :::::::::       :::   ::::::|||
     208 ValThrAla......ValProAlaSerGly..................... 215

    1935 ACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGG 1984
              :::|||   |||   :::   :::|||::::::          ::::
     216 ....AlaProAlaAlaValSerValPheGlyAlaAsnGluLeuThrValA 231

    1985 AAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGC 2034
         ::|||
     231 spGly.............................................. 232

    2035 ACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTTC 2084
                                       ||||||   ||||||:::|||:::
     233 ..................GlyHisIleThrGlyGlyArgAlaAla..... 241

    2085 CCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCC 2134
              :::||||||   :::::::|||
     242 ....GlyValAlaAlaMetAspGly........................... 248

    2135 AAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGT 2184

     248 .................................................. 248

    2185 TCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGACGA 2234

     248 .................................................. 248

    2235 TAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCC 2284
              :::|||:::   |||::::::::::   ||||||
     249 ....AlaIleValHisLeuGlnArgAlaThrIleArg.............. 259

    2285 TTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGC 2334
                                                      :::|||
     260 ..............................................ArgGly 261

    2335 AATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCAC 2384
         :::      |||||||||            :::|||      :::||||::
     262 AspAlaProAlaGlyGly............AlaValProGlyGlyAlaVa 274

    2385 CCAAAACGGC.....................................A 2395
         :   :::|||                                    :
     274 lProGlyGlyPheGlyProLeuLeuAspGlyTrpTyrGlyValAspValS 291

    2396 ACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTA 2445
         ::   |||   |||      ||||||:::      :::          |||
     291 erAspSerThrValAspLeuAlaGlnSerIleValGluAlaProGlnLeu 307

    2446 AACGGCAACACATCGGCTTCGGACAATGCTTCATTTAATCTAAGC..... 2490
         ::::::         |||:::   :::|||      :::::|||
     308 GlyAlaAlaIleArgAlaGlyArgGlyAlaArgValThrValSerGlyGl 324

    2491 .................AACAACGCCGTACAAAACGGC.......... 2511
                           :::|||:::::::::::::|||
```

```
 4073 ATTACCGATACGAAAAC...GTCAATATCGCCACCCCGGGCCTTGCATTC 4119
      ::|||||||      |||   :::::::::         |||   :::
  798 erTyrArgAlaAlaAsnGlyLeuArgValArgAspGluGlyGlySerSer 814

 4120 AACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCA 4169
             |||   |||:::::::::         ::::::   |||
  815 ValLeuGlyArgLeuGlyLeuGluValGlyLysArgIleGluLeuAlaGl 831

 4170 ACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTT 4219
           ::::::   ||||||::::::   |||      :::      :
  831 yGlyArgGlnValGlnProTyrIleLysAlaSerValLeuGlnGluPheA 848

 4220 CCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCGCAGGATTTC 4269
      ::|||       :::|||:::|||     :::        ::::::
  848 spGlyAlaGlyThrValArgThrAsnGlyIleAlaHisArgThrGluLeu 864

 4270 GGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTT 4319
           ||||||   ||||||   |||
  865 ArgGlyThrArg...AlaGluLeuGly..................... 872

 4320 CACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCG. 4368
         |||::::::   ||||||   :::::::|||   :::|||   |||
  873 ....LeuGlyMetAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrAlaS 888

 4369 .....CAGCACAGCGCGGGCATCAAATTAGGCTACCGCTGG 4404
           ::::::|||   |||   ||||||:::      |||
  888 erTyrGluTyrSerLysGlyProLysLeuAlaMetProTrp 901

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30294

seq_documentation_block:
ID   AAW30294 standard; Protein; 1477 AA.
XX
AC   AAW30294;
XX
DT   14-APR-1998  (first entry)
XX
DE   Non-typeable Haemophilus high mol.wt. surface protein HMW2.
XX
KW   Non-typeable Haemophilus; high molecular weight surface protein;
KW   HMW2; hmw2A gene; immunogen; vaccine; otitis media.
XX
OS   Haemophilus influenzae strain 12.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 34
FT                   /note= "encoded by TTC"
FT   Misc-difference 35
FT                   /note= "encoded by CGC"
FT   Misc-difference 36
FT                   /note= "encoded by TAT"
FT   Misc-difference 37
FT                   /note= "encoded by GTT"
FT   Misc-difference 38
FT                   /note= "encoded by ACT"
FT   Misc-difference 39
FT                   /note= "encoded by ATC"
FT   Misc-difference 40
FT                   /note= "encoded by TTT"
FT   Misc-difference 41
FT                   /note= "encoded by AGG"
FT   Misc-difference 42
FT                   /note= "encoded by TGT"
FT   Misc-difference 43
FT                   /note= "encoded by AAC"
FT   Misc-difference 424
FT                   /note= "encoded by TAT"
FT   Misc-difference 426
FT                   /note= "encoded by TCC"
FT   Misc-difference 428
FT                   /note= "encoded by GAC"
FT   Misc-difference 429
FT                   /note= "encoded by AGC"
FT   Misc-difference 434
FT                   /note= "encoded by AAA"
FT   Misc-difference 435
FT                   /note= "encoded by ACA"
FT   Misc-difference 442
FT                   /note= "encoded by CCT"
FT   Misc-difference 444
FT                   /note= "encoded by GAT"
FT   Misc-difference 446
FT                   /note= "encoded by ACA"
FT   Misc-difference 448
FT                   /note= "encoded by GAA"
FT   Misc-difference 454
FT                   /note= "encoded by CGC"
XX
PN   WO9736914-A1.
XX
PD   09-OCT-1997.
XX
PF   01-APR-1997;   97WO-US04707.
XX
PR   01-APR-1996;   96US-0617697.
XX
PA   (BARE/) BARENKAMP S J.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1997-503038/46.
DR   N-PSDB; AAT90995.
XX
PT   High molecular weight proteins of non-typeable Haemophilus
PT   influenzae - useful for vaccine production
XX
PS   Claim 7; Page 73-78; 183pp; English.
XX
CC   This protein comprises the high molecular weight surface protein
CC   HMW2 (123 kDa) of non-typeable Haemophilus influenzae strain 12 that
CC   has the immunological ability to protect against disease caused by a
CC   non-typeable Haemophilus strain and is characterised by at least
CC   one surface-exposed B-cell epitope that is recognised by monoclonal
CC   antibody AD6.  The HMW2 amino acid sequence was deduced from the
CC   hmw2 gene sequence (see AAT90995 and AAT90997).  The expressed protein
CC   is truncated, starting at residue 442 of the full-length gene
CC   product.  HMW1 (see AAW30293), HMW3 (see AAW30291) and HMW4 (see
CC   AAW30292) have also been identified.  A conjugate comprising HMW2
CC   linked to an antigen, hapten or polysaccharide, and a synthetic
CC   peptide of 6-150 amino acids corresponding to at least protective
CC   epitope of HMW2 are also claimed.  HMW proteins, conjugates and
CC   peptides can be used in vaccines, as immunogens for preparation of
CC   antibodies and as antigens for detection of these antibodies.
XX
SQ   Sequence   1477 AA;


alignment_scores:
            Quality:  273.50              Length:   1316
              Ratio:  0.423                 Gaps:     61
 Percent Similarity:  49.088   Percent Identity: 19.681

alignment_block:
 US-09-303-518D-653 x AAW30294   ..

 Align seg 1/1  to: AAW30294  from: 1  to: 1477

  706 AATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAAAATTAAACA 755
      ||||||      ::::::|||:::::::::::::::|||:::||||||
  165 AsnGlyPheThrAlaSerThrLeuAspIleSerAsnGluAsnIleLys.. 180

  756 TAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCT 805
                                                      :
  181 .....................................................A 181

  806 CACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGG 855
```


181 laArgAsnPheThrPheGluGlnThrLysAspLysAlaLeuAlaGlu... 196
856 GTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCT 905
197 IleValAsnHisGlyLeuIleThrValGlyLysAspGlySerValAsnLe 213
906 AGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATT 955
213 uIle..... 214
956 CAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAACGACAAT 1005
215 .....GlyGlyLysVal..... 218
1006 AATAATGGCGCAGGAAAAATCGATGCCAAACATAAACACTATTCTCTACC 1055
219 ...LysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeuLe 234
1056 TTATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCG 1105
234 uAlaGlyGlnLysIleThrIleSerAspIleIleAsnProThrIleThrT 251
1106 AGACAGCAAGAGAACCT.....GTTTAT 1128
251 yrSerIleAlaAlaProGluAsnGluAlaValAsnLeuGlyAspIlePhe 267
1129 CATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGA 1178
268 AlaLysGlyGlyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGl 284
1179 AAATATTTCC.....TTTATTGACAAAGGAAAA...GGTGAATTGATAC 1219
284 yLysLeuSerAlaAspSerValSerLysAspLysSerGlyAsnIleValL 301
1220 TTACCAGCAACATCAACCAAGGC...GCGGGCGGTTTGTATTTTGAGGGT 1266
301 euSerAlaLysGluGlyGluAlaGluIleGlyGlyVal..... 313
1267 AATTTTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGT 1316
314 .....IleSerAlaGlnAsnGlnGlnAla...LysGlyGlyLysLe 326
1317 TCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAA 1366
326 uMetIleThr...GlyAspLysValThrLeuLys...ThrGlyAlaValI 341
1367 ACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTG...GTTCAAGCC 1413
341 leAspLeuSerGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAspGlu 357
1414 AAAGGGGAAAACCAAGGCTCGGTC.....AGCGTGGG 1445
358 ArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLysLysThrSerLeuGl 374
1446 CGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAAC 1495
374 uLysGlySerThrIleAsnValSerGlyLysGluLysGlyGlyArgAlaI 391
1496 AAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTG 1545
391 leValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAlaGlnGly 407
1546 AATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGG 1595
408 SerGlyAsp.....IleAlaLysThrGl 415
1596 CGGACGTTTGGATTTGAACGGGCATTCG..... 1623
415 yGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsnAlaI 432
1624 .....CTTTCGTTCCACCGCATTCAAAATACC 1650
432 leValAspAlaLysGluTrpLeuLeuAspPheAspAsnValSerIleAsn 448
1651 GATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGT 1700
449 AlaGluAspProLeuPhe...AsnAsnThrGlyIleAsnAspGluPhePr 464
1701 TACCATTACAGGC.....AATAAAGATATTACTA 1729
464 oThrGlyThrGlyGluAlaSerAspProLysLysAsnSerGluLeuLysT 481
1730 CAACCGGCAATAACAACAACTTGGATAGCAAAAAAGAAATTGCCTACAAC 1779
481 hrThrLeuThrAsnThrThrIleSerAsnTyrLeuLys.....Asn 494
1780 GGTTGGTTTGGCGAGAAAGATGCAACCAAA.....ACGAACGGGCG 1820
495 AlaTrpThrMetAsnIleThrAlaSerArgLysLeuThrValAsnSerSe 511
1821 GCTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTT. 1869
511 rIleAsnIleGly.....SerAsnSerHisLeuIleLeuH 523
1870 .....TCCGGCGGAACAAATTTAAACGGCAATATCACG 1902
523 isSerLysGlyGlnArgGlyGlyGlyValGlnIleAspGlyAspIleThr 539
1903 CAAACAAACGGCAAACTG.....TTTTTCAGCGGCAGACCGACACCGCA 1946
540 SerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrpValAspValHi 556
1947 CGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATG..... 1983
556 sLysAsnIleThrLeuAspGlnGlyPheLeuAsnIleThrAlaAlaServ 573
1984 .....GAAGGT..... 1989
573 alAlaPheGluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLys 589
1990 ...ATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCAC 2036
590 IleValAlaGlnGlyThrValThrIleThrGlyGluGlyLysAsp..... 604
2037 ATTTAAAGCGGAAAACTTCCATATTCAGGGC..... 2067
605 .PheArgAlaAsnAsnValSerLeuAsnGlyThrGlyLysGlyLeuAsnI 621
2067 2067
621 leIleSerServalAsnAsnLeuThrHisAsnLeuSerGlyThrIleAsn 637
2068 ......GGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGA 2111
638 IleSerGlyAsnIleThrIleAsnGlnThrThrArgLysAsnThrSerTy 654
2112 TTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATC 2161
654 rTrpGlnThrSer.....HisA 660
2162 AAAGCCACACAATCTGTACACGTTCGGACTGG......ACGGGTCTGACA 2205
660 spSerHis.....TrpAsnValSerAlaLeuAsn 669
2206 AGTTGTACCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGAG 2255
670 LeuGluThrGlyAlaAsnPheThrPheIleLysTyrIleSerSerAsnSe 686
2256 CAAG.....ACCGACATCAGAGGCAATGTCAGCCTTGCCGATCACG 2296
686 rLysGlyLeuThrThrGlnTyrArgSerSer.....A 697
2297 CTCATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTTAGT... 2343
697 laGlyValAsnPheAsnGly.....ValAsnGlyAsnMetSerPhe 710


2344 .....GCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCAC 2384
711 AsnLeuLysGluGlyAlaLysValAsnPheLysLeuLysProAsnGluAs 727
2385 CCAAAAC.....GGCAACCTCAGCC 2404
727 nMetAsnThrSerLysProLeuProIleArgPheLeuAlaAsnIleThrA 744
2405 TCGTGGGCAATGCCCAAGCAACATTTAAT...CAAGCCACATTAAACGGC 2451
744 laThrGlyGlyGlySerValPhePheAspIleTyrAlaAsnHisSerGly 760
2452 AACACATCGGCTTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGT 2501
761 ArgGlyAlaGluLeuLysMetSerGluIleAsnIleSerAsnGlyAla.. 776
2502 ACAAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCA..... 2541
777 .....AsnPheThrLeuAsnSerHisValArgGlyAspAspAlaP 790
2542 .....AACGTAAGCCATTCCGCACTCAACGGCAATGTCTCC 2577
790 heLysIleAsnLysAspLeuThrIleAsnAlaThrAsnSerAsnPheSer 806
2578 CTAGCCGATAAGGCAGTATTCCATTTTGAA..... 2607
807 LeuArgGlnThrLysAspAspPheTyrAspGlyTyrAlaArgAsnAlaIl 823
2608 .AACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTAC 2656
823 eAsnSerThrTyrAsnIleSerIleLeuGlyGlyAsnValThr...LeuG 839
2657 ACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAAT 2706
839 lyGlyGlnAsnSer.....SerSerSerIleThrGlyAsn 850
2707 TTAAACCTTGACAAC...GCCACCATTACACTCAATTCCGCCTATCGACA 2753
851 IleThrIleGluLysAlaAlaAsnValThrLeu..... 861
2754 CGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCC 2803
862 .....GluAlaAsnAsnAlaProAsnGlnG 870
2804 GTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCC 2853
870 lnAsnIleArgAspArgVal.....Ile 877
2854 CGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATT 2903
878 LysLeuGlySerLeuLeuValAsnGlySerLeuSerLeuThrGlyGluAs 894
2904 CCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGG 2953
894 nAlaAspIle.....LysGlyAsnLeuThrIleS 904
2954 CGGAAAGTTCC.....GAA 2967
904 erGluSerAlaThrPheLysGlyLysThrArgAspThrLeuAsnIleThr 920
2968 GGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCT 3017
921 GlyAsnPheThr.....AsnAsnGlyThrAlaGluIleAsnIl 933
3018 CGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAA 3067
933 eThrGlnGlyValValLysLeuGly.....AsnValThrAsnAspGlyA 948
3068 ATCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGT 3117
948 spLeuAsnIleThrThrHisAlaLysArgAsnGlnArgSerIleIleGly 964


3118 TATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAA 3167
965 GlyAspIleIleAsnLysLysGlySerLeuAsnIleThrAspSerAsnAs 981
3168 AGAACAAGAGCTT..... 3180
981 nAspAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysGluGlyAsnL 998
3181 .....TCCGACAAACTC..... 3192
998 euThrIleSerSerAspLysIleAsnIleThrLysGlnIleThrIleLys 1014
3193 ...GGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACA 3239
1015 LysGlyIleAspGlyGluAspSerSerSerAspAlaThrSerAsnAlaAs 1031
3240 ACTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACG 3289
1031 nLeuThrIleLysThrLysGluLeuLys.....LeuThrGluA 1044
3290 CGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTGCC 3339
1044 spLeuSerIleSerGlyPheAsn.....LysAlaGluIleThrAla 1057
3340 GAACCGGCCCGGCAG.....GCAGGCGGGGAAAATGCCGGCATTATGCA 3383
1058 LysAspGlyArgAspLeuThrIleGlyAsnSerAsnAspGlyAsnSerGl 1074
3384 GGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGG 3433
1074 yAlaGluAlaLysThrValThrPheAsnAsnValLysAspGlnLysIleS 1091
3434 CGAAACAG.....CGCGAAGCGGAAACCCGGCCGGCTACC 3468
1091 erAlaAspGlyHisAsnValThrLeuAsnSerLysValLysThrSerSer 1107
3469 ACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCA 3518
1108 SerAsnGlyGlyArgGluSerAsnSerAspAsnAspThrGlyLeuThrIl 1124
3519 GCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGC...CGTTATG 3565
1124 eThrAlaLysAsnValGluValAsnLysAspIleThrSerLeuLysThrV 1141
3566 CCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCC 3615
1141 alAsnIleThrAlaSerGluLysValThrThr.ThrAlaGlySerThrIl 1157
3616 GTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCCGT 3665
1157 eAsnAlaThrAsnGlyLysAlaSerIleThrThrLysThrGlyAspIleS 1174
3666 TTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGC...AAGATT 3712
1174 erGlyThrIleSerGlyAsnThrValSerValSerAlaThrGlyAspLeu 1190
3713 TCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAA 3762
1191 ThrThrLysSerGlySerLysIleGluAlaLysSerGlyGluAlaAsnVa 1207
3763 AACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGG 3812
1207 lThrSerAlaThrGlyThrIleGlyGly.....ThrIleSerGlyA 1221
3813 AAACACCTTCGACGACGGCATCGGCA..... 3838
1221 snThrValAsnValThrAlaAsnAlaGlyAspLeuThrValGlyAsnGly 1237
3839 .....ACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAA 3876
1238 AlaGluIleAsnAlaThrGluGlyAlaAlaThrLeuThrAlaThrGlyAs 1254
3877 T.....ACGGCATCGGCAGGTTCGACATCGGCATCAGCGCGGGCGCGGG 3920

```
      |      |||:::::::|||||||||:::   :::|||:::::::::::::|
  1254 nThrLeuThrThrGluAlaGlySerSerIleThrSerThrLysGlyGlnV 1271

  3921 TTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGAGGCAAAATCCGCCGCC 3970
       ||   :::   |||      |||::::::|||   |||   |||
  1271 alAspLeuLeuAla......GlnAsnGlySerIleAlaGlySerIleAsn 1285

  3971 GCGTGCTGCATTACGGCATTCAGGCAAGATACCGCGCAGGTTTCGGCGGA 4020
       |||:::   :::|||               |||      :::   |||
  1286 AlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrThrValAlaGl 1302

  4021 TTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAA 4066
        |||      :::   |||||||:::      :::|||:::::::|||
  1302 ySerAspIleLysAlaThrSerGlyThrLeuValIleAsnAlaLys 1317

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01827

seq_documentation_block:
ID   AAB01827 standard; Protein; 969 AA.
XX
AC   AAB01827;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain Joyc mature HMW2A protein, SEQ ID NO:32.
XX
KW   Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain Joyc.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52178.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 8; Fig 19A-O; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   969 AA;


alignment_scores:
          Quality:  273.00              Length:   1164
            Ratio:  0.487                 Gaps:     54
 Percent Similarity:  48.110   Percent Identity: 20.619

alignment_block:
 US-09-303-518D-653 x AAB01827   ..

 Align seg 1/1  to: AAB01827  from: 1  to: 969

  1090 AATGTTTCTTTATCC.................GAGACAGCAAGAGAACC 1121
       |||||||||::::::                  :::   ::::::|||
     3 AsnValSerIleGluAsnProSerThrGluArgAsnAspSerAsnGluAs 19

  1122 TGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATA 1171
        :::   :::::::::   |||   |||      :::|||::::::|||| 
    19 pLeuGluTyrThrGlyThrGlyGluAsnIleAsnAsnProLysValAsnA 36

  1172 ATGGA.............................GAAAATATT..... 1185
       ||                                 |||||||| 
    36 snGlnSerLysLysThrLeuThrSerSerIleLeuGluAsnIleLeuLys 52

  1186 ......TCCTTTATTGAC...AAAGGAAAAGGTGAATTGATACTTACCAG 1226
             ||||||::::::      :::      ::::::   ::::::|| 
    53 LysGlySerPheValAsnIleThrAlaThrAspAsnIleTyrValAsnSe 69

  1227 CAACATCAACCAAGGC...GCGGGCGGTTTGTATTTTGAGGGTAATTTTA 1273
       |:::||||||   |||   :::|||   |||   :::::::|||:::
    69 rSerIleAsnIleGlyAspSerGlyHisLeuIleLeuSerGlyGlyGlyA 86

  1274 CGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATC 1323
                                      :::|||:::||||||   |||
    86 rg..........................AsnGlyGlyGlyValLysIle 93

  1324 AGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCG 1373
       :::   :::   |||   |||
    94 AsnLysAsnIleThrSerThr............................. 100

  1374 CCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGG.... 1419
                          |||:::|||   :::::::::||||||
   101 ...............GlyGlySerLeuThrIleAsnSerLysGlyTrpV 112

  1420 ..GAAAACCAAGGCTCGGTCAGCGTGGGC..................... 1446
         :::   ::::::::::::|||:::|||
   112 alAspIleHisSerAsnIleSerLeuGlyThrGlyPheLeuAsnIleThr 128

  1447 ...GACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAA 1493
         :::|||:::|||   :::::::::          |||::::::::::
   129 SerAsnGlySerValAlaPheGluLysAlaAspLysAspLysAlaArgSe 145

  1494 ACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAAC 1543
       :   |||   :::::|||           :::::::|||   ::::::|
   145 rAlaAlaAspAlaGlnIle.........ValAlaGlnGlyIleIleAsnL 159

  1544 TGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGC 1593
       ||:::::::::|||:::                                |||||
   159 euThrGlyGluAsnLys.....................ThrPheArg 167

  1594 GGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCA 1643
         ::::::::::::|||||||||   :::   ::::::         ::
```

```
 168 LeuAsnAsnValSerLeuAsnGlyValGlyGlnGlyLeuSerIleThrSe 184

1644 AAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAAT 1693
     :|||·                 :::   |||:::::::::          ::::
 184 rAsn.................ValGlyAsnGlnThrHisLysPheAspG 195

1694 CCACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGC...... 1737
     ::   ::::::||||||||||||      |||:::   |||:::
 195 lyGluIleAsnIleThrGlyAsnValThrIleAsnGlnThrAlaProAla 211

1738 .....................................AATAACAACAA 1748
                                          |||   :::::
 212 ThrThrAlaTyrTrpAsnPheSerTyrAspSerTyrTrpAsnValSerTh 228

1749 CTTGGATAGCAAAAAAGAAATTGCCTAC.....................A 1777
     :|||:::   :::|||:::   ::::::                     :
 228 rLeuAsnValGlnLysAsnSerSerPheThrPheIleLysArgThrGluS 245

1778 ACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCTCAAT 1827
     :::::   ||||||          ::::::::::|||   :::
 245 erAsnArgPheGlyProThrThrProLeuArgSerSerGlyGlyValPhe 261

1828 CTGAATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGG 1877
     :::|||                                          ||
 262 PheAsn.........................................Gl 264

1878 AACAAATTTAAACGGCAATATC.........ACGCAAACAAACGGCAAAC 1918
     ||||      ||||||||||:::          :::   ||||||:::::::
 264 yThr......AsnGlyAsnMetValLeuAsnValGlyThrAsnSerArgV 279

1919 TGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGC 1968
     :::::|||:::   :::|||:::   ::::::   |||
 279 alLeuPheAsnLeuLysProAsnGluAsnThrAsnAsn............ 291

1969 GGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAA 2018
          ||||||
 292 ......SerLysPro.................................. 294

2019 CGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCCATATTCAGGGCG 2068
           :::         |||:::|||:::          |||   ||||
 295 .......LeuProLeuGlnPheAsnAlaAsnIleThrAlaIleGlyGlyG 309

2069 GACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCAT 2118
     ||                          :::|||:::   |||   |||
 309 ly.........................SerValSerPheAspIleHis 316

2119 TTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCA 2168
        :::||||||:::           |||   :::   :::   :::
 317 ...AlaAsnHisSer............GlyArgGlyAlaGluLeuLysMe 328

2169 CACAATCTGTACACGTTCGGACTGGACGGGTCTGACA......AGTTGTA 2212
     :::     :::   ||||||   |||:::||||||      |||   :
 328 tAsnThrIleAsnIleSerAspGlyThrSerLeuThrLeuGlnSerHisV 345

2213 CCGAAAAA............ACCATTACCGACGATAAAGTGATTGCTTCA 2250
     :::::|||              |||:::   |||   :::|||   ::::
 345 alArgLysAspSerAlaPheIleIleSerLysAspLeuThrIleAsnAla 361

2251 TTG..............................AGCAAGACCGACAT 2267
                                       |||   ||||||
 362 ThrGlySerAsnPheThrLeuGluGlnSerProAspSerPheThrAspLy 378

2268 CAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGAC 2317
           :::   ::::::::::::      ::::::::::::::::|||
 378 sTyrProGlyArgAlaIleSerSerThrLysAsnIleThrIleSerGly. 394

2318 TTGCCACACTCAACGGCAATCTTAGTGCAGGCGGAGACACGCACTATACG 2367
                  ||||||:::|||   ||||||
 395 ..............GlyAsnValSerLeuGlyGly.............. 401

2368 GTTACGCGCAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAATGC 2417
                       ||||||::::::   ::::::   ||||||
 402 .................GlnAsnSerSerSerAspIleLysGlyAsnIl 412

2418 CCAA...GCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTT 2464
           :::   ::::::::::||||||:::::::::::|||
 412 eThrIleLysSerSerThrAsnValThrLeuLysAlaHisAsnSerProA 429

2465 CGGACAATGCTTCATTTAATCTAAGC...AACAACGCCGTACAAAACGGC 2511
       |||   ||||||   :::|||:::   :::|||         :::|||
 429 rgAspPheAlaSerArgThrLeuThrLeuGlyAsnLeuAsnValGluGly 445

2512 AGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACT 2561
     :::|||||||||:::                          |||   :::   ::
 446 AsnLeuThrLeuThrGly.....................SerValAlaAspIl 456

2562 CAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT.........T 2602
     ::::||||||:::|||:::   :::   |||:::|||
 456 eLysGlyAsnLeuSerIleLeuAsnAspAlaThrPheLysGlyGluThrS 473

2603 TTGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCA 2652
       ||||||   ::::::||||||:::::::::::::::::      :::
 473 erGluAsnLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaAsp 489

2653 TTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGC...ACGGAATT 2699
     ::::::::::|||:::::::      ::::::   ::::::   |||:::
 490 IleAsnIleLysGlnGlyValValAsnIleGlnGlyAsnIleThrAsnLy 506

2700 AGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATC 2749
       |||:::||||||:::
 506 sGlyGlyLeuAsnIle................................ 511

2750 GACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGC 2799

 511 ..................................................... 511

2800 CGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGA 2849
                                                 |||:::|||::
 512 ..................................ThrThrAsnAlaGl 516

2850 ATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAA 2899
     :::::::   :::|||:::   :::||||||:::::::::::::::|||
 516 nAsnAsnGlnLysThrIle...IleAsnGlyAsnIleThrAsnGluGly. 531

2900 CATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAG 2949
                                          |||   |||:::
 532 .....................................GlyAspLeuAsn 535

2950 CTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTCAACAAT..... 2994
     :::   :::|||:::::::::::          :::|||
 536 IleLysAspSerAsnAsnAsnAlaGluIleGlnIleGlyGlyAsnIleSe 552

2995 .......ACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGG 3037
              ||||||      :::|||   :::::::::::::::::::
 552 rGlnLysLysGlyAsnLeuThrIleSerSerAspLysIleAsnIleThrL 569

3038 AA..............................GGAAAAGACAACACA 3054
     ::                                |||:::|||::::::
 569 ysLysIleThrIleLysAlaGlyValAspGluGlyGlySerAspSerSer 585

3055 CCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATGC 3104
     |||   |||   |||   |||:::|||::::::
 586 ProAlaSer...AsnAlaAsnLeuThrIleLys................. 595

3105 CGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGC 3154

 595 ..................................................... 595

3155 ATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCGGGA 3204
            :::|||      ||||||::      |||:::   :::|||
 596 ........ThrLysThrLeuGluLeuThrGlyAspLeuAsnIleSerGly 609

3205 GAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCCAAACA 3254
        ::::::|||   :::|||||||||:::::::::      :::
 610 PheAsnLysAlaGluIleThrAlaLysAsnGlyAsnAspLeuThrIleGl 626
```

```
3255 ACAGGCGGAAAAAGACAACGCGCAAAGC..................... 3282
      :::|||:::      ||||||::::::
 626 yLysAlaSerAspGlyAsnAlaAsnAlaLysLysValThrPheAspLysV 643

3283 ..CTTGACGCGCTGATTGCGGCC...GGGCGCAATGCCACC......GAA 3321
        |||:::   |||:::|||   |||:::|||:::|||      :::
 643 alLysAspSerLysIleSerAlaAsnGlyHisAsnValThrLeuAsnSer 659

3322 AAGGCAGAA......AGTGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGA 3365
     |||:::|||      :::      :::   |||   :::::::::   ::
 660 LysValGluThrSerAsnSerAspSerSerAlaAspAspSerAsnAspAs 676

3366 AAATGCCGGCATTATGCAGGCGGAGGAAGAG................... 3396
     :|||:::|||:::      :::   ::::::
 676 nAsnThrGlyLeuThrIleSerAlaLysAspValThrValAsnAsnAspV 693

3397 ........AAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAA 3438
               |||   ::::::   :::      |||:::   :::
 693 alThrSerHisLysThrIleAsnIleSerAlaThrThrGlyAsnValThr 709

3439 CAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCG 3488
        :::|||:::   |||      |||:::|||::: |||        :
 710 ThrLysGluSerThrThrIleAsnAlaAlaThrGly.SerValGluValT 726

3489 CCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAAC 3538
     ::|||   :::|||         :::         :::::::::   |||
 726 hrAlaLysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValAsn 742

3539 CGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTT 3588
        :::||||||   |||   :::         :::
 743 ValThrAlaThrAspSerLeuThrThrGlnAlaSer.............. 754

3589 TCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGT 3638
               |||   :::   |||      :::|||:::   ||||||
 755 .......SerSerIleThrSerSerAsnGlyGlnThrThrLeuThrAlaL 769

3639 GTTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACA 3688
           :::   |||:::::         :::|||||||:::   |||
 769 ysAsnGlySerIleAlaGlySerIle......AspAlaAlaAsnValThr 783

3689 CCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGAC 3738
        ||||||                        |||:::::      ||
 784 LeuAsnThr.......................ThrGlyThrLeuThrTh 792

3739 CTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCG...TCGG 3785
     |   |||   |||      :::   ||||||:::::   |||   :::|
 792 rValAlaGlySerAsnIleLysAlaThrSerGlyThrLeuAlaIleAsnA 809

3786 CATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACGGCATCG 3835
     ||:::      :::      |||                 |||||||||
 809 laLysAspAlaLysLeuAspGly.................ThrAlaSer 819

3836 GCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCATC 3885
     :::   |||            ||||||      :::   |||::::::::
 820 GlyAspArg............ThrValValAsnAlaThrAsnAlaSerGl 832

3886 GGCAGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAG 3935
     ::::||||||::::::|||::::::   ::::::|||   :::::::::
 832 ySerGlySerValThrAlaAlaThrSerSerAsnValAsnIleThrGlyA 849

3936 CCTTTCAGACG.....................GCATCAGAGGCAAAATCC 3964
       :::::::|||                    :::::::::::|||:::
 849 spLeuSerThrIleAsnGlyLeuAsnIleIleSerLysAsnGlyLysAsn 865

3965 GCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCGCGCAGGTTTC 4014
     :::::::::         :::   :::                |||   ::
 866 ThrValValLeuLysGlyAlaGluIleAspValLysTyrIleGlnProGl 882

4015 GGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCA 4064
     ::::   ||||||::::::      :::|||:::||||||:::::::::
 882 yValAlaSerAlaAsnGluValIleGluAlaLysArgAlaLeuGluLysV 899

4065 AAAAGCGGATTACCGAT.................ACGAAAACGTCAATA 4096
```

```
        |||   :::::::|||                  :::|||   ::::::
 899 alLysAspLeuSerAspGluGluArgGluThrLeuAlaLysLeuGlyVal 915

4097 TCGCCACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCA 4146
     |||      |||   ::::::   :::   |||
 916 SerAlaValArgPheValGluProAsnAsnThr.................. 926

4147 GATTATTCATTCAAACCGGCGCAACACATTTCCATCACGC 4186
          |||      |||   :::|||   |||   :::|||
 927 ....IleThrValAsnThrGlnAsnGluPheThrThrArg 938
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01826

seq_documentation_block:

ID AAB01826 standard; Protein; 975 AA.
XX
AC AAB01826;
XX
DT 11-SEP-2000 (first entry)
XX
DE Haemophilus influenzae strain Joyc HMW2A protein, SEQ ID NO:30.
XX
KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
OS Haemophilus influenzae strain Joyc.
XX
PN WO200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
DR WPI; 2000-303789/26.
DR N-PSDB; AAA52177.
XX
PT Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
PS Claim 12; Fig 19A-O; 307pp; English.
XX
CC The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMWA
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,

```
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   975 AA;


alignment_scores:
           Quality:  273.00            Length:   1164
             Ratio:  0.487              Gaps:     54
 Percent Similarity:  48.110   Percent Identity: 20.619

alignment_block:
 US-09-303-518D-653 x AAB01826   ..

 Align seg 1/1  to: AAB01826  from: 1  to: 975

   1090 AATGTTTCTTTATCC.................GAGACAGCAAGAGAACC 1121
        |||||||||:::::                  :::   ::::::|||
      9 AsnValSerIleGluAsnProSerThrGluArgAsnAspSerAsnGluAs 25

   1122 TGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATA 1171
         :::   :::::::::   |||   |||      :::|||::::::||||
     25 pLeuGluTyrThrGlyThrGlyGluAsnIleAsnAsnProLysValAsnA 42

   1172 ATGGA............................GAAAATATT..... 1185
        ||                              |||||||||
     42 snGlnSerLysLysThrLeuThrSerSerIleLeuGluAsnIleLeuLys 58

   1186 ......TCCTTTATTGAC...AAAGGAAAAGGTGAATTGATACTTACCAG 1226
              ||||||::::::      :::      ::::::   ::::::||
     59 LysGlySerPheValAsnIleThrAlaThrAspAsnIleTyrValAsnSe 75

   1227 CAACATCAACCAAGGC...GCGGGCGGTTTGTATTTTGAGGGTAATTTTA 1273
        |:::||||||   |||   :::|||   |||   ::::::|||:::
     75 rSerIleAsnIleGlyAspSerGlyHisLeuIleLeuSerGlyGlyGlyA 92

   1274 CGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATC 1323
                                   :::|||:::||||||   |||
     92 rg.........................AsnGlyGlyGlyValLysIle 99

   1324 AGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCG 1373
        :::   :::   |||   |||
    100 AsnLysAsnIleThrSerThr............................ 106

   1374 CCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGG.... 1419
                        |||:::|||   :::::::::||||||
    107 ...............GlyGlySerLeuThrIleAsnSerLysGlyTrpV 118

   1420 ..GAAAACCAAGGCTCGGTCAGCGTGGGC..................... 1446
          :::   ::::::::::::|||:::|||
    118 alAspIleHisSerAsnIleSerLeuGlyThrGlyPheLeuAsnIleThr 134

   1447 ...GACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAA 1493
           :::|||:::|||   :::::::::         |||::::::::::
    135 SerAsnGlySerValAlaPheGluLysAlaAspLysAspLysAlaArgSe 151

   1494 ACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAAC 1543
        :   |||   :::::|||                :::::|||   ::::::|
    151 rAlaAlaAspAlaGlnIle.........ValAlaGlnGlyIleIleAsnL 165

   1544 TGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGC 1593
        ||:::::::::|||:::                             ||||||
    165 euThrGlyGluAsnLys.......................ThrPheArg 173

   1594 GGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCA 1643
           ::::::::::::|||||||||   :::   ::::::         ::
    174 LeuAsnAsnValSerLeuAsnGlyValGlyGlnGlyLeuSerIleThrSe 190

   1644 AAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAAT 1693
        :|||                  :::   |||:::::::::      ::::

    190 rAsn.................ValGlyAsnGlnThrHisLysPheAspG 201

   1694 CCACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGC...... 1737
        ::   ::::::||||||||||||      |||:::   |||:::
    201 lyGluIleAsnIleThrGlyAsnValThrIleAsnGlnThrAlaProAla 217

   1738 .....................................AATAACAACAA 1748
                                             |||   :::::
    218 ThrThrAlaTyrTrpAsnPheSerTyrAspSerTyrTrpAsnValSerTh 234

   1749 CTTGGATAGCAAAAAAGAAATTGCCTAC.....................A 1777
        :|||:::   :::|||:::   ::::::                    :
    234 rLeuAsnValGlnLysAsnSerSerPheThrPheIleLysArgThrGluS 251

   1778 ACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCTCAAT 1827
        :::::   ||||||               ::::::::::|||   :::
    251 erAsnArgPheGlyProThrThrProLeuArgSerSerGlyGlyValPhe 267

   1828 CTGAATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGG 1877
        :::|||                                         ||
    268 PheAsn.........................................Gl 270

   1878 AACAAATTTAAACGGCAATATC.........ACGCAAACAAACGGCAAAC 1918
        ||||      |||||||||::          :::   ||||||::::::
    270 yThr......AsnGlyAsnMetValLeuAsnValGlyThrAsnSerArgV 285

   1919 TGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGC 1968
        :::::|||:::   :::|||:::   ::::::   |||
    285 alLeuPheAsnLeuLysProAsnGluAsnThrAsnAsn............ 297

   1969 GGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAA 2018
              ||||||
    298 ......SerLysPro................................... 300

   2019 CGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCCATATTCAGGGCG 2068
              :::          |||:::|||:::          |||   ||||
    301 .......LeuProLeuGlnPheAsnAlaAsnIleThrAlaIleGlyGlyG 315

   2069 GACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCAT 2118
        ||                           :::|||:::   |||   |||
    315 ly.........................SerValSerPheAspIleHis 322

   2119 TTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCA 2168
           :::||||||:::             |||   :::   :::   :::
    323 ...AlaAsnHisSer............GlyArgGlyAlaGluLeuLysMe 334

   2169 CACAATCTGTACACGTTCGGACTGGACGGGTCTGACA......AGTTGTA 2212
         :::      :::   ||||||   |||:::||||||      |||   :
    334 tAsnThrIleAsnIleSerAspGlyThrSerLeuThrLeuGlnSerHisV 351

   2213 CCGAAAAA............ACCATTACCGACGATAAAGTGATTGCTTCA 2250
        ::::::|||                |||:::   |||   :::|||   :::
    351 alArgLysAspSerAlaPheIleIleSerLysAspLeuThrIleAsnAla 367

   2251 TTG................................AGCAAGACCGACAT 2267
                                           |||   ||||||
    368 ThrGlySerAsnPheThrLeuGluGlnSerProAspSerPheThrAspLy 384

   2268 CAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGAC 2317
           :::   ::::::::::::      :::::::::::::::|||
    384 sTyrProGlyArgAlaIleSerSerThrLysAsnIleThrIleSerGly. 400

   2318 TTGCCACACTCAACGGCAATCTTAGTGCAGGCGGAGACACGCACTATACG 2367
                  ||||||:::|||   ||||||
    401 .............GlyAsnValSerLeuGlyGly.............. 407

   2368 GTTACGCGCAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAATGC 2417
                        ||||||::::::   ::::::   ||||||
    408 .................GlnAsnSerSerSerAspIleLysGlyAsnIl 418

   2418 CCAA...GCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTT 2464
              :::   :::::::::||||||:::::::::::|||
    418 eThrIleLysSerSerThrAsnValThrLeuLysAlaHisAsnSerProA 435
```

```
2465 CGGACAATGCTTCATTTAATCTAAGC...AACAACGCCGTACAAAACGGC 2511
      |||   ||||||   :::|||:::   :::|||          :::|||
 435 rgAspPheAlaSerArgThrLeuThrLeuGlyAsnLeuAsnValGluGly 451

2512 AGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACT 2561
     :::||||||||||:::                    |||   :::   ::
 452 AsnLeuThrLeuThrGly.................SerValAlaAspIl 462

2562 CAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT.........T 2602
     ::::||||||:::|||:::   :::   |||:::|||
 462 eLysGlyAsnLeuSerIleLeuAsnAspAlaThrPheLysGlyGluThrS 479

2603 TTGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCA 2652
       ||||||   ::::::||||||:::::::::::::::      :::
 479 erGluAsnLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaAsp 495

2653 TTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGC...ACGGAATT 2699
     :::::::::|||::::::     ::::::  ::::::   |||:::
 496 IleAsnIleLysGlnGlyValValAsnIleGlnGlyAsnIleThrAsnLy 512

2700 AGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATC 2749
      |||:::||||||:::
 512 sGlyGlyLeuAsnIle................................. 517

2750 GACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGC 2799

 517 ................................................... 517

2800 CGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGA 2849
                                          |||:::|||::
 518 ...................................ThrThrAsnAlaGl 522

2850 ATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAA 2899
     :::::::   :::|||:::   :::||||||:::::::::::::|||
 522 nAsnAsnGlnLysThrIle...IleAsnGlyAsnIleThrAsnGluGly. 537

2900 CATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAG 2949
                                             |||   |||:::
 538 .....................................GlyAspLeuAsn 541

2950 CTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTCAACAAT..... 2994
     :::   :::|||:::::::::::::          :::|||
 542 IleLysAspSerAsnAsnAsnAlaGluIleGlnIleGlyGlyAsnIleSe 558

2995 .......ACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGG 3037
             ||||||      :::|||   ::::::::::::::::::::
 558 rGlnLysLysGlyAsnLeuThrIleSerSerAspLysIleAsnIleThrL 575

3038 AA..............................GGAAAAGACAACACA 3054
     ::                              |||:::|||::::::
 575 ysLysIleThrIleLysAlaGlyValAspGluGlyGlySerAspSerSer 591

3055 CCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATGC 3104
     |||   |||   |||   |||:::|||::::::
 592 ProAlaSer...AsnAlaAsnLeuThrIleLys.................. 601

3105 CGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGC 3154

 601 ................................................... 601

3155 ATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCGGGA 3204
             :::|||      ||||||:::      |||:::   :::|||
 602 ........ThrLysThrLeuGluLeuThrGlyAspLeuAsnIleSerGly 615

3205 GAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCCAAACA 3254
       :::::::|||   :::||||||||||:::::::::      :::
 616 PheAsnLysAlaGluIleThrAlaLysAsnGlyAsnAspLeuThrIleGl 632

3255 ACAGGCGGAAAAAGACAACGCGCAAAGC..................... 3282
      :::|||:::      ||||||::::::
 632 yLysAlaSerAspGlyAsnAlaAsnAlaLysLysValThrPheAspLysV 649

3283 ..CTTGACGCGCTGATTGCGGCC...GGGCGCAATGCCACC......GAA 3321
       |||:::   |||:::|||   |||:::|||:::|||      :::
 649 alLysAspSerLysIleSerAlaAsnGlyHisAsnValThrLeuAsnSer 665

3322 AAGGCAGAA......AGTGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGA 3365
     |||:::|||      :::      :::   |||   :::::::::   ::
 666 LysValGluThrSerAsnSerAspSerSerAlaAspAspSerAsnAspAs 682

3366 AAATGCCGGCATTATGCAGGCGGAGGAAGAG.................. 3396
     :|||:::|||:::      :::   ::::::
 682 nAsnThrGlyLeuThrIleSerAlaLysAspValThrValAsnAsnAspV 699

3397 ........AAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAA 3438
               |||   ::::::   :::      |||:::   :::
 699 alThrSerHisLysThrIleAsnIleSerAlaThrThrGlyAsnValThr 715

3439 CAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCG 3488
       :::|||:::   |||      |||:::|||::: |||        :
 716 ThrLysGluSerThrThrIleAsnAlaAlaThrGly.SerValGluValT 732

3489 CCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAAC 3538
     ::|||   :::|||         :::         :::::::::   |||
 732 hrAlaLysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValAsn 748

3539 CGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTT 3588
       :::||||||   |||   :::         :::
 749 ValThrAlaThrAspSerLeuThrThrGlnAlaSer.............. 760

3589 TCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGT 3638
               |||   ::   |||      :::|||:::   ||||||
 761 .......SerSerIleThrSerSerAsnGlyGlnThrThrLeuThrAlaL 775

3639 GTTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACA 3688
             :::   |||:::::         :::|||||:::   |||
 775 ysAsnGlySerIleAlaGlySerIle......AspAlaAlaAsnValThr 789

3689 CCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGAC 3738
       ||||||                        |||::::::      ||
 790 LeuAsnThr.......................ThrGlyThrLeuThrTh 798

3739 CTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCG...TCGG 3785
     |   |||   |||      :::   ||||||::::::   |||   :::|
 798 rValAlaGlySerAsnIleLysAlaThrSerGlyThrLeuAlaIleAsnA 815

3786 CATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACGGCATCG 3835
     ||:::      :::      |||                 |||||||||
 815 laLysAspAlaLysLeuAspGly.................ThrAlaSer 825

3836 GCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCATC 3885
     :::   |||             ||||||      :::   |||::::::::
 826 GlyAspArg............ThrValValAsnAlaThrAsnAlaSerGl 838

3886 GGCAGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAG 3935
     ::::||||||:::::::|||:::::::   ::::::|||   :::::::::
 838 ySerGlySerValThrAlaAlaThrSerSerAsnValAsnIleThrGlyA 855

3936 CCTTTCAGACG.....................GCATCAGAGGCAAAATCC 3964
       :::::|||                     :::::::::::|||:::
 855 spLeuSerThrIleAsnGlyLeuAsnIleIleSerLysAsnGlyLysAsn 871

3965 GCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCGCGCAGGTTTC 4014
     :::::::::          :::   :::                |||   ::
 872 ThrValValLeuLysGlyAlaGluIleAspValLysTyrIleGlnProGl 888

4015 GGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCA 4064
     ::::   ||||||::::::      :::|||:::||||||::::::::::
 888 yValAlaSerAlaAsnGluValIleGluAlaLysArgAlaLeuGluLysV 905

4065 AAAAGCGGATTACCGAT.................ACGAAAACGTCAATA 4096
       |||   ::::::|||                 :::|||   ::::::
 905 alLysAspLeuSerAspGluGluArgGluThrLeuAlaLysLeuGlyVal 921

4097 TCGCCACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCA 4146
```

```
      |||      |||   ::::::   :::   |||
  922 SerAlaValArgPheValGluProAsnAsnThr..... 932

 4147 GATTATTCATTCAAACCGGCGCAACACATTTCCATCACGC 4186
      |||      |||   :::|||   |||   :::|||
  933 ....IleThrValAsnThrGlnAsnGluPheThrThrArg 944
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01841

seq_documentation_block:

ID AAB01841 standard; Protein; 1004 AA.
XX
AC AAB01841;
XX
DT 11-SEP-2000 (first entry)
XX
DE Haemophilus influenzae strain PMH1 mature HMW2A protein, SEQ ID NO:57.
XX
KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
OS Haemophilus influenzae strain PMH1.
XX
PN WO200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
DR WPI; 2000-303789/26.
DR N-PSDB; AAA52190.
XX
PT Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
PS Claim 8; Fig 25A-O; 307pp; English.
XX
CC The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMWA
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typeable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents a mature HMWA protein from a non-typeable
CC strain of H. influenzae.
XX
SQ Sequence 1004 AA;

alignment_scores:

```
            Quality:  272.00              Length:   1116
              Ratio:  0.514                 Gaps:     49
 Percent Similarity:  47.401   Percent Identity: 20.878
```

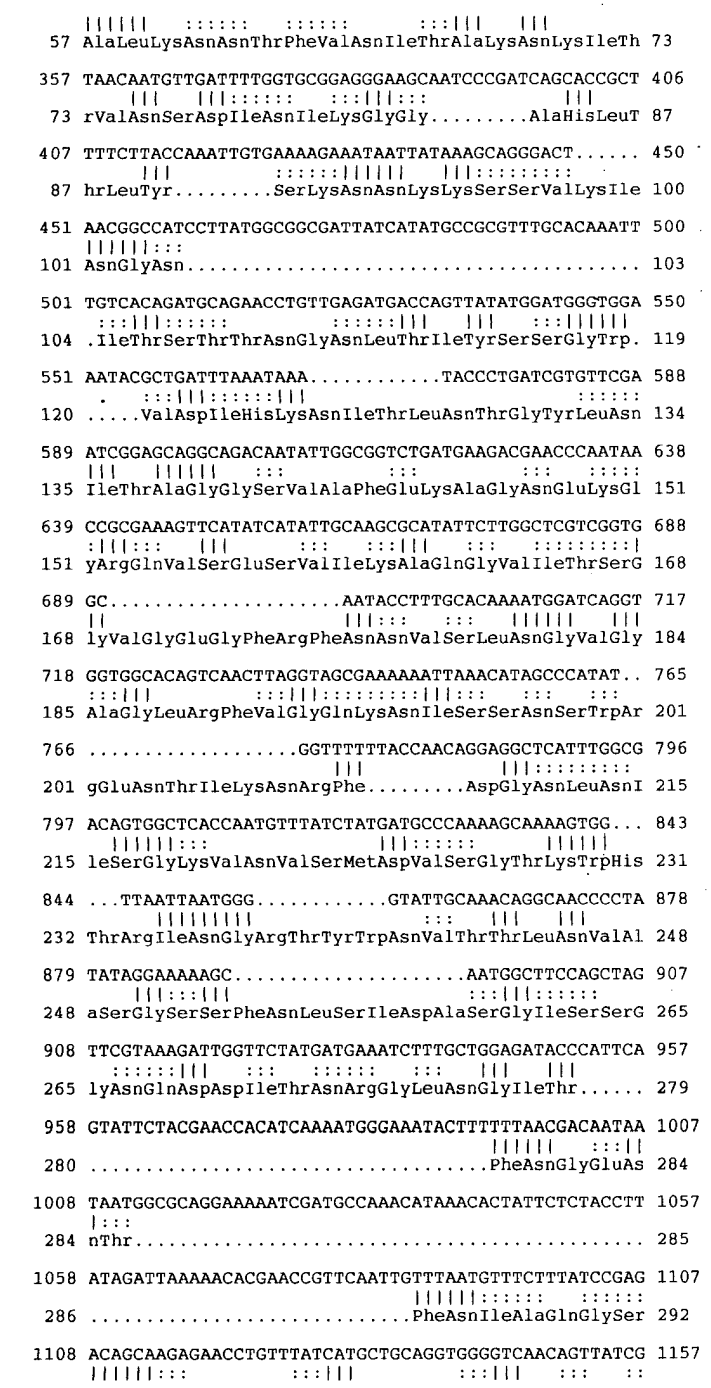
alignment_block:

US-09-303-518D-653 x AAB01841 ..

Align seg 1/1 to: AAB01841 from: 1 to: 1004

```
  307 GCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTA 356
      ||||||   ::::::   ::::::      :::|||   |||
   51 AlaLeuLysAsnAsnThrPheValAsnIleThrAlaLysAsnLysIleTh 67

  357 TAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCGATCAGCACCGCT 406
        |||   |||::::::   :::|||:::            |||
   67 rValAsnSerAspIleAsnIleLysGlyGly.....AlaHisLeuT 81

  407 TTTCTTACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACT..... 450
        |||         ::::::||||||   |||:::::::::
   81 hrLeuTyr.....SerLysAsnAsnLysLysSerSerValLysIle 94

  451 AACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATT 500
      ||||||:::
   95 AsnGlyAsn..................................... 97

  501 TGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATGGGTGGA 550
       :::|||:::::         :::::::|||   |||   :::||||||
   98 .IleThrSerThrThrAsnGlyAsnLeuThrIleTyrSerSerGlyTrp. 113

  551 AATACGCTGATTTAAATAAA...........TACCCTGATCGTGTTCGA 588
        :::|||:::::::|||                          ::::::
  114 .....ValAspIleHisLysAsnIleThrLeuAsnThrGlyTyrLeuAsn 128

  589 ATCGGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAACCCAATAA 638
      |||   ||||||   :::         :::         :::   :::::
  129 IleThrAlaGlyGlySerValAlaPheGluLysAlaGlyAsnGluLysGl 145

  639 CCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTG 688
      :|||:::   |||       :::   :::|||   :::   :::::::::|
  145 yArgGlnValSerGluSerValIleLysAlaGlnGlyValIleThrSerG 162

  689 GC.....................AATACCTTTGCACAAAATGGATCAGGT 717
      ||                     |||:::   :::   ||||||   |||
  162 lyValGlyGluGlyPheArgPheAsnAsnValSerLeuAsnGlyValGly 178

  718 GGTGGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATAT.. 765
      :::|||         :::|||:::::::::|||:::   :::   :::
  179 AlaGlyLeuArgPheValGlyGlnLysAsnIleSerSerAsnSerTrpAr 195

  766 ..................GGTTTTTTACCAACAGGAGGCTCATTTGGCG 796
                         |||          |||:::::::::
  195 gGluAsnThrIleLysAsnArgPhe.....AspGlyAsnLeuAsnI 209

  797 ACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGG... 843
        ||||||:::               |||::::::      ||||||
  209 leSerGlyLysValAsnValSerMetAspValSerGlyThrLysTrpHis 225

  844 ...TTAATTAATGGG............GTATTGCAAACAGGCAACCCCTA 878
         |||||||||              :::   |||   |||
  226 ThrArgIleAsnGlyArgThrTyrTrpAsnValThrThrLeuAsnValAl 242

  879 TATAGGAAAAAGC..................AATGGCTTCCAGCTAG 907
        |||:::|||                    :::|||::::::
  242 aSerGlySerSerPheAsnLeuSerIleAspAlaSerGlyIleSerSerG 259

  908 TTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCA 957
      :::::::|||   :::   ::::::   :::   |||   |||
```

```
 2626 AAAATCAGCGGC.........GGCAAGGATACGGCATTACACTTAAAAGA 2666
        |||||||||         ||||||   :::::::::      :::
  740 AspIleSerGlyThrIleSerGlyLysThrValSerValThrAlaSerTh 756

 2667 CAGCGAATGGACGCTGCCGTCGGGC........................ 2691
       ::::::   |||:::   ::::::
  756 rGlyAspLeuThrValArgLysAlaAlaThrIleSerValThrGluGlyA 773

 2692 ..........ACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATT 2730
                 |||      ||||||             :::|||
  773 laAlaThrLeuThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySer 789

 2731 ACACTCAATTCCGCCTATCGACACGATGCGGCAGGCGCGCAAACCGGCAG 2780
      ::::::::::|||:::      :::         :::||||||   ||||||
  790 SerIleThrSerThrLysGlyGlnValAspLeuSerAlaGlnAspGlySe 806

 2781 TGCGGCAGAT...........GCGCCGCGCCGCCGTTCGCGCCGTTCCC 2818
      |   |||               |||            :::      :::
  806 rIleAlaGlyGlnIleSerAlaAlaAsnValThrLeuAsnThrThrGlyT 823

 2819 TATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTG 2868
        |||:::::::::      ::::::   ::::::      :::||||||
  823 hrLeuThrThrValGluGlySerAsnIleLysAlaThrSerGlyThrLeu 839

 2869 ACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGA 2918
      ::::::|||:::|||
  840 AlaIleAsnAlaLys..................................... 844

 2919 ACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAAG 2968
                      ::::::||||||      :::   |||:::::
  845 ................AspAlaLysLeuAspGlyThrAlaSerGlyAsnA 856

 2969 GCACTTACACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCTC 3018
        |||   :::   |||:::|||   :::|||:::      |||:::
  856 rgThrGluValAsnAlaThrAsnAlaSerGlySerGlySerValThrAla 872

 3019 GAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAA 3068
      :::      :::   |||:::        :::                :::::
  873 LysThrSerSerAsnValAsnIleThrGlyAspLeuSerThrIleAsnGl 889

 3069 TCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTT 3118
      :||||||:::      :::|||               |||
  889 yLeuAsnIleIleSerGluAsn..............GlyArgAsnThrV 901

 3119 ATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCG...... 3162
        :::|||      |||:::   :::   :::      :::|||
  901 alArgLeuArgGlyLysGluIleAspValLysTyrIleGlnProGlyVal 917

 3163 :.....................................GTCAAAGAACA 3173
                                            |||  ::
  918 AlaSerValGluGluValIleGluAlaLysArgValLeuGluLysValLy 934

 3174 AGAGCTTTCCGACAAA............CTCGGCAAGGCGGGAGAAACAG 3211
      ::::|||||||||:::           |||:::|||   |||   :::
  934 sAspLeuSerAspGluGluArgGluThrLeuAlaLysLeuGlyValSerA 951

 3212 AGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCG 3261
        :::   ::::::      :::      ::::::   :::   |||
  951 laValArgPheValGluProAsnAsnAlaIleThrIleAsnThrGlnAsn 967

 3262 GAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGC 3309
      |||      :::      |||   :::::::::|||:::   |||:::
  968 GluPheThrThrArgProSerSerGlnValIleIleSerGluGlyLys 983

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01840

seq_documentation_block:
ID   AAB01840 standard; Protein; 1010 AA.
XX
AC   AAB01840;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain PMH1 HMW2A protein, SEQ ID NO:55.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain PMH1.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52189.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 25A-O; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1010 AA;


alignment_scores:
            Quality:  272.00              Length:   1116
              Ratio:  0.514                 Gaps:     49
 Percent Similarity:  47.401   Percent Identity: 20.878

alignment_block:
 US-09-303-518D-653 x AAB01840   ..

 Align seg 1/1  to: AAB01840  from: 1  to: 1010

     307 GCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTA 356
```


57 AlaLeuLysAsnAsnThrPheValAsnIleThrAlaLysAsnLysIleTh 73
357 TAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCGATCAGCACCGCT 406
73 rValAsnSerAspIleAsnIleLysGlyGly.....AlaHisLeuT 87
407 TTTCTTACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACT..... 450
87 hrLeuTyr.....SerLysAsnAsnLysLysSerSerValLysIle 100
451 AACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATT 500
101 AsnGlyAsn..... 103
501 TGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATGGGTGGA 550
104 .IleThrSerThrThrAsnGlyAsnLeuThrIleTyrSerSerGlyTrp. 119
551 AATACGCTGATTTAAATAAA.....TACCCTGATCGTGTTCGA 588
120 .....ValAspIleHisLysAsnIleThrLeuAsnThrGlyTyrLeuAsn 134
589 ATCGGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAACCCAATAA 638
135 IleThrAlaGlyGlySerValAlaPheGluLysAlaGlyAsnGluLysGl 151
639 CCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTG 688
151 yArgGlnValSerGluSerValIleLysAlaGlnGlyValIleThrSerG 168
689 GC.....AATACCTTTGCACAAAATGGATCAGGT 717
168 lyValGlyGluGlyPheArgPheAsnAsnValSerLeuAsnGlyValGly 184
718 GGTGGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATAT.. 765
185 AlaGlyLeuArgPheValGlyGlnLysAsnIleSerSerAsnSerTrpAr 201
766 .....GGTTTTTTACCAACAGGAGGCTCATTTGGCG 796
201 gGluAsnThrIleLysAsnArgPhe.....AspGlyAsnLeuAsnI 215
797 ACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGG... 843
215 leSerGlyLysValAsnValSerMetAspValSerGlyThrLysTrpHis 231
844 ...TTAATTAATGGG.....GTATTGCAAACAGGCAACCCCTA 878
232 ThrArgIleAsnGlyArgThrTyrTrpAsnValThrThrLeuAsnValAl 248
879 TATAGGAAAAAGC.....AATGGCTTCCAGCTAG 907
248 aSerGlySerSerPheAsnLeuSerIleAspAlaSerGlyIleSerSerG 265
908 TTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCA 957
265 lyAsnGlnAspAspIleThrAsnArgGlyLeuAsnGlyIleThr..... 279
958 GTATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAACGACAATAA 1007
280 .....PheAsnGlyGluAs 284
1008 TAATGGCGCAGGAAAAATCGATGCCAAACATAAACACTATTCTCTACCTT 1057
284 nThr..... 285
1058 ATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAG 1107
286 .....PheAsnIleAlaGlnGlySer 292
1108 ACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCG 1157
293 ThrAlaAsn.....PheHisIleLysThrSerValMetThrProLy 306
1158 ACCCAGACTGAAT.....AATGGAGAAAATATTTCCTTTATTG 1195
306 sProAsnSerAsnTyrAlaLeuPheAsnGly...AsnIleSerValLeuG 322
1196 ACAAAGGA.....AAAGGTGAATTGATACTTACCAGCAACATCAACCAA 1239
322 lyGlyGlyThrValAsnPheGluLeuAsnAlaSerSerSerThrHisThr 338
1240 GGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCCTAAAAA 1289
339 ThrSerGlyAlaIleIleAsnSerGlnAsnPhe..... 349
1290 CAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGGCAGTACCG 1339
350 .....AsnValSerGlyGlySerLysL 357
1340 TTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGC 1389
357 euAsnLeuLysAlaSerGlySerThrAsnThrAlaPhe.....LeuIle 371
1390 AAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAAC..... 1425
372 LysAsnAsnLeuThrLeuAsnAlaThrGlyGlyAsnIleGluIleLysGl 388
1426 ....CAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGC 1471
388 nValGluGlyThrAspSerArgIleGlnLysGlyValValAlaGluGlnA 405
1472 AGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTC 1521
405 snIleIlePheGluGly.....GlyAsnIleThrLeuGly 416
1522 AGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGA 1571
417 SerGlnLysAlaProThrGluIleLysGlyAspValThrValLys..... 431
1572 CAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATT 1621
432 .....GlnGlyThrAsnAlaThrLeuArgSerAlaA 442
1622 CGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTC 1671
442 snPheAspAsnHisLys.....GlyAlaLeuIleVal 452
1672 AACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGA 1721
453 AsnGlyAsn.....ValThrAlaAsnGlyAs 461
1722 TATTACTACAACCGGCAAT.....AACAACAACTTGGATAGCA 1759
461 nLeuThrAlaAspGlyAspThrIleLysIleLysGlyAsnLeuAspValA 478
1760 AAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAA 1809
478 laGlnGlyAlaLysPheAsnGly.....SerThrLys 488
1810 ACGAACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCAC 1859
489 AsnAsnLeuAsnIleThrGlyThrPheThrAsnAsnGlyThrSerIleIl 505
1860 TTTACTGCTTTCCGGCGGAACAAATTTAAACGGCAATATCACGCAAACAA 1909
505 eAspIleThrGlnGlyValValAsnLeu...GlyAsnValThr...AsnA 520
1910 ACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCAT 1959
520 spGlyLysLeu.....AsnIleThrThrHisAlaLys..... 530
1960 TTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATCGT 2009
531 .....SerGlyGlnLysSerIleIleArgGlyAspIleIl 542

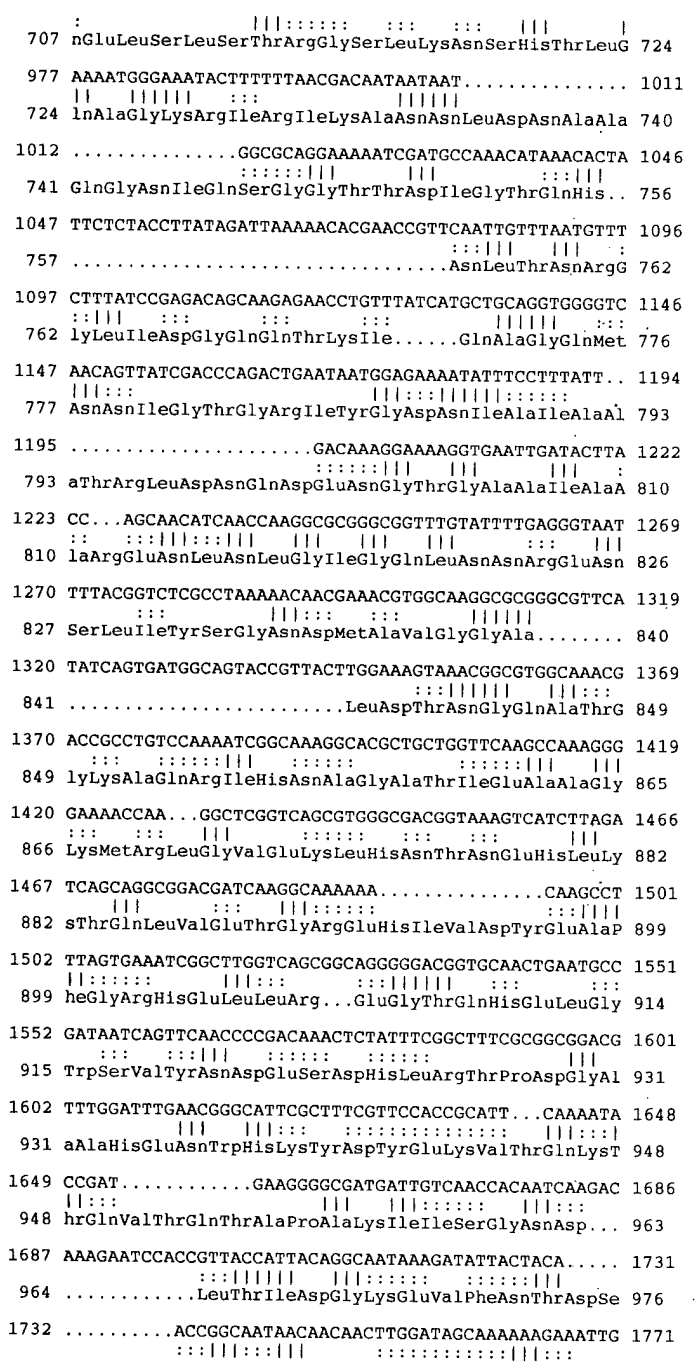
```
2010 GTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCCATA 2059
542 eAsnLysGlnGlyAsnLeuAsnIleThrAspAsnAsnSerAsnAlaGluI 559

2060 TTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGC 2109
559 leGlu..............IleGlyGlyAsnIleSerGlnLysGluGly 570

2110 GATTGGCATTTAAGCAATCACGCCCAAGCAGTT..................  2142
571 AsnLeuThrIleSerSerAspLysValAsnIleThrLysGlnIleThrIl 587

2143 ....TTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGG 2188
587 eLysAlaGlyVal...AspGlyGluSerSerSerSerSerThrAlaSerA 603

2189 ACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGACGATAAA 2238
603 spAlaAsnLeuThrIleLysThrLysGluLeuThrPheThrAspAsnLeu 619

2239 GTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGC 2288
620 AsnIleSerGlyPheAsnLysAlaGluIle............ThrAlaLy 632

2289 CGATCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGCAAT. 2337
632 sAspAsnSerAspLeuIleIleGlyLysAlaSerSerAspAsnSerAsnA 649

2338 .....CTTAGTGCAGGC 2349
649 laLysGlnValThrPheAspLysValLysAspSerLysIleSerAlaGly 665

2350 GGAGACACGCACTATACGGTTACGCGCAACGCC.........ACCCAAAA 2390
666 .........AsnHisAsnValThrLeuAsnSerLysValGluThrSerAs 679

2391 C..............GGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAA 2425
679 nSerAspGlySerThrGlyAsnGlySerAspAspAsnAsnIleGlyLeuT 696

2426 CATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCT 2475
696 hrIleSerAlaLysAspValThrValAsnSerAsnIleThrSerHisLys 712

2476 TCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTC 2525
713 ThrValAsnIleSer.........AlaSerGluGlyGlyIleThr..... 724

2526 CGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCT 2575
725 .......ThrLysAlaGlyThrThrIleAsnAlaThrThrGlySerValG 739

2576 CCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGA 2625
739 luValThrAlaLys............................ThrGly 745

2626 AAAATCAGCGGC.........GGCAAGGATACGGCATTACACTTAAAAGA 2666
746 AspIleSerGlyThrIleSerGlyLysThrValSerValThrAlaSerTh 762

2667 CAGCGAATGGACGCTGCCGTCGGGC.......................... 2691
762 rGlyAspLeuThrValArgLysAlaAlaThrIleSerValThrGluGlyA 779

2692 ..........ACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATT 2730
779 laAlaThrLeuThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySer 795

2731 ACACTCAATTCCGCCTATCGACACGATGCGGCAGGCGCGCAAACCGGCAG 2780
796 SerIleThrSerThrLysGlyGlnValAspLeuSerAlaGlnAspGlySe 812

2781 TGCGGCAGAT............GCGCCGCGCCGCCGTTCGCGCCGTTCCC 2818
812 rIleAlaGlyGlnIleSerAlaAlaAsnValThrLeuAsnThrThrGlyT 829

2819 TATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTG 2868
829 hrLeuThrThrValGluGlySerAsnIleLysAlaThrSerGlyThrLeu 845

2869 ACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGA 2918
846 AlaIleAsnAlaLys..................................  850

2919 ACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAAG 2968
851 ................AspAlaLysLeuAspGlyThrAlaSerGlyAsnA 862

2969 GCACTTACACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCTC 3018
862 rgThrGluValAsnAlaThrAsnAlaSerGlySerGlySerValThrAla 878

3019 GAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAA 3068
879 LysThrSerSerAsnValAsnIleThrGlyAspLeuSerThrIleAsnGl 895

3069 TCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTT 3118
895 yLeuAsnIleIleSerGluAsn..............GlyArgAsnThrV 907

3119 ATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCG...... 3162
907 alArgLeuArgGlyLysGluIleAspValLysTyrIleGlnProGlyVal 923

3163 .....GTCAAAGAACA 3173
924 AlaSerValGluGluValIleGluAlaLysArgValLeuGluLysValLy 940

3174 AGAGCTTTCCGACAAA............CTCGGCAAGGCGGGAGAAACAG 3211
940 sAspLeuSerAspGluGluArgGluThrLeuAlaLysLeuGlyValSerA 957

3212 AGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCG 3261
957 laValArgPheValGluProAsnAsnAlaIleThrIleAsnThrGlnAsn 973

3262 GAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGC 3309
974 GluPheThrThrArgProSerSerGlnValIleIleSerGluGlyLys 989
```

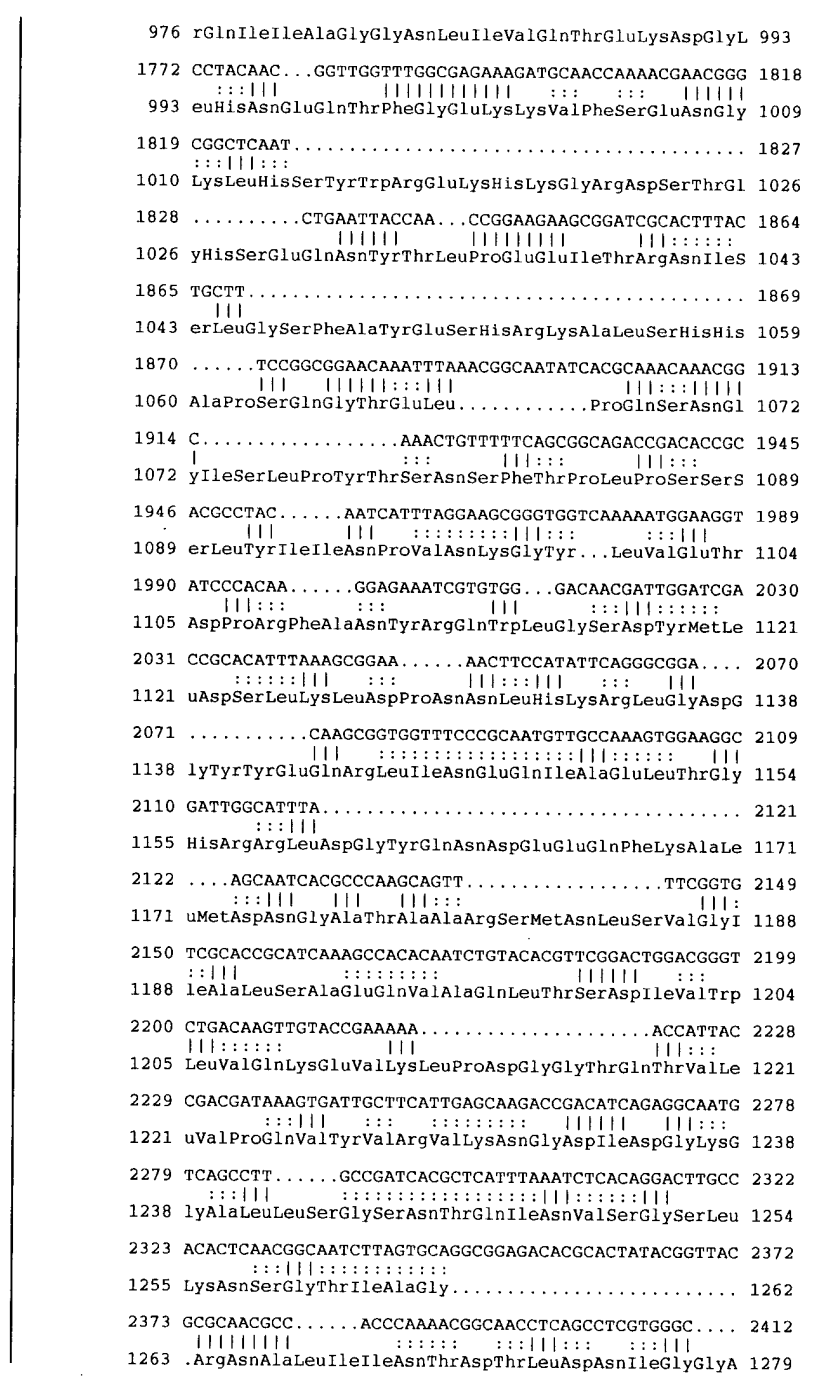
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75097

seq_documentation_block:

ID AAY75097 standard; Protein; 2514 AA.
XX
AC AAY75097;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 564 protein sequence SEQ ID NO:1668.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.

```
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53859.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 854-855; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   2514 AA;


alignment_scores:
            Quality:  271.00              Length:   1854
              Ratio:  0.328                 Gaps:     91
 Percent Similarity:  44.606   Percent Identity: 19.471

alignment_block:
 US-09-303-518D-653 x AAY75097   ..

 Align seg 1/1  to: AAY75097  from: 1  to: 2514

      6 AACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCTAAAACCG 55
        |||   ::::::|||||||||
    364 AsnAlaLysAspThrAspAsn..............................  370

     56 GCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGC 105
                                    :::|||       :::|||
    371 .....................ThrAlaGluHisLysValAsnIleArgS  380

    106 ATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACCA 155
          ::::::   :::   :::||||||   :::   :::::: :::   ||
    380 erGlnGlyValGluAsnSerGlyThrAlaValSerGlnGln.GlyThrGl 396

    156 ATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGA 205
        |   :::   :::      :::|||   |||   :::   :::::::::
    396 nIleHisSerGlnSerIleGlnAsnThrGlyThrLeu...LeuSerSerG 412

    206 AAGATATTGAGGTTTACAACAAA.........AAAGGGGAGTTGGTCGGC 246
          :::|||   ::::::|||:::         |||:::|||      |||
    412 lyGluIleLeuIleHisAsnSerGlySerLeuLysAsnGluThrSerGly 428

    247 AAATCGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAA 296
                  :::            ||||||   ::::::::::::::::
    429 ThrIleGluAlaAlaArgLeuAlaIleAspThrAspThrLeuAsnAsnGl 445

    297 CGGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATA 346
        :|||      :::   :::|||:::|||   :::
    445 nGlyLysLeuSerGlnThrGlySerGlnLysLeu.................  456

    347 ACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCGAT 396
                      ::::::|||      |||:::|||:::      :::
    457 ..............HisIleAsp......AlaGlnGlyLysMetAspAsn 466

    397 CAGCACCGCTTTTCTTACCAA......ATTGTGAAAAGAAATAATTATAA 440
        :::   |||:::::   |||                    :::|||   ::
    467 ArgGlyArgMetGlyLeuGlnAspThrAlaProThrAlaSerAsnGlySe 483

    441 AGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCG.... 486
        :::::::   :::|||:::   |||:::::::::::::|||
    483 rSerAsnGlnThrGlyAsnSerTyrAsnAlaSerPheHisSerSerThrT 500

    486 ..................................................  486

    500 hrThrProThrThrAlaThrGlyThrGlyThrAlaThrValSerIleSer 516

    487 .................................CGTTTGCACAAATT 500
                                        |||   |||
    517 AsnIleThrAlaProThrPheAlaAspGlyThrIleArgThrHisGlyAl 533

    501 TGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATGGGTGGA 550
          :::   ::::::      :::      ::::::     |||
    533 aLeuAspAsnSerGlySerIleIleAlaAsnGlyGlnThrAspValSerA 550

    551 AATACGCTGATTTAAATAAA.............................. 570
                |||||||:::
    550 laGlnGlnGlyLeuAsnAsnAlaGlyGlnIleAspIleHisGlnLeuAsn 566

    571 ....................................TACCCTGATCG 581
                                            |||
    567 AlaLysGlySerAlaPheAspAsnHisAsnGlyThrIleIleSerAspAl 583

    582 TGTTCGAATCGGAGCAGGC.............................. 600
          |||:::|||   ||||||
    583 aValHisIleGlnAlaGlySerLeuAsnAsnGlnAsnGlyAsnIleThrT 600

    601 ..AGACAATATTGGCGGTCTGATGAAGACGAACCCAATAACCGCGAAAGT 648
          ||||||      :::   :::   |||:::   :::|||   ::::::
    600 hrArgGlnGlnLeuGluIleGluThrAspGlnLeuAspAsnAlaHisGly 616

    649 TCATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAATACCTT 698
        :::          |||:::      :::      |||:::|||:::
    617 LysLeuLeuSerAlaGluIleAlaAspLeuAlaValSerGlySerLeuAs 633

    699 TGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAAAA 748
            |||||||||:::   :::
    633 nAsnGlnAsnGlyGluIleAla.............................  640

    749 TTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGAC 798

    640 ..................................................  640

    799 AGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAAT 848
        :::::::::   ::::::|||:::|||:::|||::::::      :::||
    641 ThrAsnGlnGlnLeuIleIleHisAspGlyGlnGlnSerThrAlaValIl 657

    849 T.........AATGGGGTATTGCAAACAGGC................... 870
        |         ||||||::::::|||:::|||
    657 eAspAsnThrAsnGlyThrIleGlnSerGlyArgAspValAlaIleGlnA 674

    871 .................AACCCCTATATAGGAAAAAGCAATGGCTTCCAG 903
                        |||      ::::::   :::|||   ::::::
    674 laLysSerLeuSerAsnAsnGlyThrLeuAlaAlaAspAsnLysLeuAsp 690

    904 CTAGTTCGTAAAGATTGGTTCTATGATGAA......ATCTTTGCTGGAGA 947
        ::::::   :::|||   ||||||   |||      |||   ||||||::
    691 IleAlaLeuGlnAspAspPheTyrValGluArgAsnIleValAlaGlyAs 707

    948 T..............ACCCATTCAGTATTCTACGAACCACAT......C 976
```


707 nGluLeuSerLeuSerThrArgGlySerLeuLysAsnSerHisThrLeuG 724
977 AAAATGGGAAATACTTTTTTAACGACAATAATAAT..... 1011
724 lnAlaGlyLysArgIleArgIleLysAlaAsnAsnLeuAspAsnAlaAla 740
1012 .....GGCGCAGGAAAAATCGATGCCAAACATAAACACTA 1046
741 GlnGlyAsnIleGlnSerGlyGlyThrThrAspIleGlyThrGlnHis.. 756
1047 TTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTT 1096
757 .....AsnLeuThrAsnArgG 762
1097 CTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTC 1146
762 lyLeuIleAspGlyGlnGlnThrLysIle.....GlnAlaGlyGlnMet 776
1147 AACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATT.. 1194
777 AsnAsnIleGlyThrGlyArgIleTyrGlyAspAsnIleAlaIleAlaAl 793
1195 .....GACAAAGGAAAAGGTGAATTGATACTTA 1222
793 aThrArgLeuAspAsnGlnAspGluAsnGlyThrGlyAlaAlaIleAlaA 810
1223 CC...AGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAAT 1269
810 laArgGluAsnLeuAsnLeuGlyIleGlyGlnLeuAsnAsnArgGluAsn 826
1270 TTTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCA 1319
827 SerLeuIleTyrSerGlyAsnAspMetAlaValGlyGlyAla..... 840
1320 TATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACG 1369
841 .....LeuAspThrAsnGlyGlnAlaThrG 849
1370 ACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGG 1419
849 lyLysAlaGlnArgIleHisAsnAlaGlyAlaThrIleGluAlaAlaGly 865
1420 GAAAACCAA...GGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGA 1466
866 LysMetArgLeuGlyValGluLysLeuHisAsnThrAsnGluHisLeuLy 882
1467 TCAGCAGGCGGACGATCAAGGCAAAAAA.....CAAGCCT 1501
882 sThrGlnLeuValGluThrGlyArgGluHisIleValAspTyrGluAlaP 899
1502 TTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCC 1551
899 heGlyArgHisGluLeuLeuArg...GluGlyThrGlnHisGluLeuGly 914
1552 GATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACG 1601
915 TrpSerValTyrAsnAspGluSerAspHisLeuArgThrProAspGlyAl 931
1602 TTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATT...CAAAATA 1648
931 aAlaHisGluAsnTrpHisLysTyrAspTyrGluLysValThrGlnLysT 948
1649 CCGAT.....GAAGGGGCGATGATTGTCAACCACAATCAAGAC 1686
948 hrGlnValThrGlnThrAlaProAlaLysIleIleSerGlyAsnAsp... 963
1687 AAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACTACA..... 1731
964 .....LeuThrIleAspGlyLysGluValPheAsnThrAspSe 976
1732 .....ACCGGCAATAACAACAACTTGGATAGCAAAAAAGAAATTG 1771


976 rGlnIleIleAlaGlyGlyAsnLeuIleValGlnThrGluLysAspGlyL 993
1772 CCTACAAC...GGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGG 1818
993 euHisAsnGluGlnThrPheGlyGluLysLysValPheSerGluAsnGly 1009
1819 CGGCTCAAT..... 1827
1010 LysLeuHisSerTyrTrpArgGluLysHisLysGlyArgAspSerThrGl 1026
1828 .....CTGAATTACCAA...CCGGAAGAAGCGGATCGCACTTTAC 1864
1026 yHisSerGluGlnAsnTyrThrLeuProGluGluIleThrArgAsnIleS 1043
1865 TGCTT..... 1869
1043 erLeuGlySerPheAlaTyrGluSerHisArgLysAlaLeuSerHisHis 1059
1870 ......TCCGGCGGAACAAATTTAAACGGCAATATCACGCAAACAAACGG 1913
1060 AlaProSerGlnGlyThrGluLeu.....ProGlnSerAsnGl 1072
1914 C.....AAACTGTTTTTCAGCGGCAGACCGACACCGC 1945
1072 yIleSerLeuProTyrThrSerAsnSerPheThrProLeuProSerSerS 1089
1946 ACGCCTAC......AATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGT 1989
1089 erLeuTyrIleIleAsnProValAsnLysGlyTyr...LeuValGluThr 1104
1990 ATCCCACAA......GGAGAAATCGTGTGG...GACAACGATTGGATCGA 2030
1105 AspProArgPheAlaAsnTyrArgGlnTrpLeuGlySerAspTyrMetLe 1121
2031 CCGCACATTTAAAGCGGAA......AACTTCCATATTCAGGGCGGA.... 2070
1121 uAspSerLeuLysLeuAspProAsnAsnLeuHisLysArgLeuGlyAspG 1138
2071 .....CAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGC 2109
1138 lyTyrTyrGluGlnArgLeuIleAsnGluGlnIleAlaGluLeuThrGly 1154
2110 GATTGGCATTTA..... 2121
1155 HisArgArgLeuAspGlyTyrGlnAsnAspGluGluGlnPheLysAlaLe 1171
2122 ....AGCAATCACGCCCAAGCAGTT.....TTCGGTG 2149
1171 uMetAspAsnGlyAlaThrAlaAlaArgSerMetAsnLeuSerValGlyI 1188
2150 TCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGT 2199
1188 leAlaLeuSerAlaGluGlnValAlaGlnLeuThrSerAspIleValTrp 1204
2200 CTGACAAGTTGTACCGAAAAA.....ACCATTAC 2228
1205 LeuValGlnLysGluValLysLeuProAspGlyGlyThrGlnThrValLe 1221
2229 CGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATG 2278
1221 uValProGlnValTyrValArgValLysAsnGlyAspIleAspGlyLysG 1238
2279 TCAGCCTT......GCCGATCACGCTCATTTAAATCTCACAGGACTTGCC 2322
1238 lyAlaLeuLeuSerGlySerAsnThrGlnIleAsnValSerGlySerLeu 1254
2323 ACACTCAACGGCAATCTTAGTGCAGGCGGAGACACGCACTATACGGTTAC 2372
1255 LysAsnSerGlyThrIleAlaGly..... 1262
2373 GCGCAACGCC......ACCCAAAACGGCAACCTCAGCCTCGTGGGC.... 2412
1263 .ArgAsnAlaLeuIleIleAsnThrAspThrLeuAspAsnIleGlyGlyA 1279

```
2413 .....AATGCCCAAGCAACATTTAATCAAGCCACA..... 2442
1279 rgIleHisAlaGlnLysSerAlaValThrAlaThrGlnAspIleAsnAsn 1295

2443 TTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAATCTAAGCAA 2492
1296 IleGlyGlyMetLeuSerAlaGluGlnThrLeuLeuLeuAsnAlaGlyAs 1312

2493 CAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAA 2542
1312 nAsnIleAsnSerGlnSerThrThrAlaSerSerGlnAsnThrGln..... 1327

2543 ACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCA 2592
1328 .....GlySerSerThrTyrLeuAspArgMet 1336

2593 GTATTCCATTTTGAAAACAGCCGCTTTACCGGAAAAATCAGCGGC..... 2637
1337 AlaGlyIleTyr.....IleThrGlyLysGluLysGlyValLe 1349

2638 .....GGCAAGGATACGGCATTA..... 2655
1349 uAlaAlaGlnAlaGlyLysAspIleAsnIleIleAlaGlyGlnIleSerA 1366

2656 ..CACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGC 2703
1366 snGlnSerGluGlnGlyGlnThrArgLeuGlnAlaGlyArgAsp..... 1380

2704 AATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACA 2753
1381 ...IleAsnLeuAspThrValGlnThrSerLysHisGlnAlaThrHisPh 1396

2754 CGATGCGGCAGGC.....GCGCAAACCGGCAGTGCGGCAGATG 2791
1396 eAspAlaAspAsnHisValIleArgGlySerThrAsnGluValGlySerS 1413

2792 CGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACT 2841
1413 erIleGlnThrLysGlyAspValThrLeuLeuSer..... 1424

2842 TCGGCAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGG 2891
1425 .....GlyAsnAsnLeuAsnAl 1430

2892 TCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCA 2941
1430 a..... 1430

2942 AATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTCAAC 2991
1431 .....LysAlaAlaGluValSerSerAlaAsnGlyThrLeuAlaValSer 1445

2992 AATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGG 3041
1446 .....Al 1446

3042 AAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACG 3091
1446 aLysAsnAspIleAsnIleSerAlaGlyIleAsnThrThr..... 1459

3092 AACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGC 3141
1460 ..HisValAspAspAlaSer.....LysHisThrGlyArgSerGlyGly 1473

3142 GAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACT 3191
1474 GlyAsnLysLeu.....ValIleThrAspLysAl 1483

3192 CGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACG.....GCAAAACAGG 3235
1483 aGlnSerHisHisGluThrAlaGlnSerSerThrPheGluGlyLysGlnV 1500

3236 CACAACTTGCCGCCAAACAACAGGCGGAA..... 3264
1500 alValLeuGlnAlaGlyAsnAspAlaAsnIleLeuGlySerAsnValIle 1516

3265 AAAGACAACGCGCAAAGCCTTGACGCG.....CTGATTGCGGCCGGGCG 3308
1517 SerAspAsnGlyThrGlnIleGlnAlaGlyAsnHisValArgIleGlyTh 1533

3309 CAATGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCAG 3358
1533 rThrGlnThrGlnSerGlnSerGluThrTyrHisGlnThrGlnLysSerG 1550

3359 GCGGGGAAAATGCCGGCATT.....ATGCAGGCGGAGGAA 3393
1550 lyLeuMetSerAlaGlyIleGlyPheThrIleGlySerLysThrAsnThr 1566

3394 GAGAAAAAACGGGTGCAGGCGGATAAA..... 3420
1567 GlnGluAsnGlnSerGlnSerAsnGluHisThrGlySerThrValGlySe 1583

3421 .....GACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGC 3460
1583 rLeuLysGlyAspThrThrIleValAlaGlyLysHisTyrGluGlnIleG 1600

3461 CGGCTACCACCGCCTTCCCCCGCGCC..... 3486
1600 lySerThrValSerSerProGluGlyAsnAsnThrIleTyrAlaGlnSer 1616

3487 .....CGCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACC 3527
1617 IleAspIleGlnAlaAlaHisAsnLysLeuAsnSerAsnThrThrGlnTh 1633

3528 GCAACCCCAA.....CCGCAGCGCGACC 3550
1633 rTyrGluGlnLysGlyLeuThrValAlaPheSerSerProValThrAspL 1650

3551 TGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTC 3600
1650 euAlaGlnGln.....AlaIleAlaValAlaGln 1659

3601 AACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGA 3650
1660 SerSerLysGlnValGlyGlnSerLysAsnAspArgValAsnAlaMetAl 1676

3651 CCGCCGCAACGCCGTTTGG..... 3669
1676 aAlaAlaAsnAlaGlyTrpGlnAlaTyrGlnThrGlyLysSerAlaGlnA 1693

3670 .....ACAAGCGGCATCCGGGACACCAAACAC..... 3696
1693 snLeuAlaAsnGlyThrThrAsnAlaLysGlnValSerIleSerIleThr 1709

3697 TACCGTTCGCAAGATTTCCGCGCCTACCGCCAA.....CAAAC 3734
1710 TyrGlyGluGlnGlnAsnArgGlnThrThrGlnValGlnAlaAsnGlnAl 1726

3735 CGACCTGCGCCAAATC..... 3750
1726 aGlnAlaSerGlnIleGlnAlaGlyGlyLysThrThrLeuIleAlaThrG 1743

3751 .....GGTATGCAGAAAAACCTCGGC.....AGCGGG 3777
1743 lyAlaAlaGluGlnSerAsnIleAsnIleAlaGlySerAspValAlaGly 1759

3778 CGCGTCGGC...ATCCTGTTTTCGCACAAC..... 3804
1760 LysAlaGlyThrIleLeuIleAlaAspAsnAspIleThrLeuGlnSerAl 1776

3805 .CGGACCGGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTG 3853
1776 aGluGlnSerAsnThr...GluArgGlyGlnAsnLysSerAlaGlyTrpA 1792

3854 CCCACGGTGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATCGGC 3903
```

```
          ||||||:::      :::::::|||   |||   ::::::::::|||
1792 snAlaGlyAlaAlaVal...SerPheGlyGlnGlyGlyTrpSerLeuGly 1807

3904 ATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAG 3953
     ::::::||||||::::::   :::::::|||             |||
1808 ValThrAlaGlyGlyAsnValGlyLysGlyTyr.........GlyAsnGl 1821

3954 AGGCAAAATCCGCCGCCGCGTGCTGCATTAC.........GGCATTCAGG 3994
       :::|||   :::|||      |||            |||   |||:
1821 yAspSerIleThrHisArgHisSerHisIleGlyAspLysGlySerGlnT 1838

3995 CAAGATACCGCGCAGGT.........................TTCGGC 4017
     ::      ::::::|||
1838 hrLeuIleGlnSerGlyGlyAspThrThrIleLysGlyAlaGlnValArg 1854

4018 GGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAA.. 4065
     |||   |||::::::      :::::::|||   :::      :::|||
1855 GlyLysGlyValGln......ValAsnAlaLysAsnLeuSerIleGlnSe 1869

4066 ..........AAAGCGGATTACCGATACGAAAACGTCAATATCGCC.... 4101
                :::      |||:::   ::::::   |||   :::
1869 rValGlnAspArgGluThrTyrGlnSerLysGlnGlnAsnAlaSerAlaG 1886

4102 .....ACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCA 4146
          |||   |||   :::|||:::          ||||||      :::
1886 lnValThrValGlyTyrGlyPheSer.........AlaGly......Gly 1897

4147 GATTATTCATTCAAACCG......GCGCAACACATTTCCATCACGCCTTA 4190
     |||||||||   :::         |||:::|||:::|||:::|||
1898 AspTyrSerGlnSerLysIleArgAlaAspHisValSerValThrGluGl 1914

4191 TTTGAGC..................................... 4197
        :::
1914 nSerGlyIleTyrAlaGlyGluAspGlyTyrGlnIleLysValGlyAsnH 1931

4198 ......................CTGTCCTATACCGATGCCGCTTCCGGC 4224
                           ::::::   |||:::::::|||:::
1931 isThrAspLeuLysGlyGlyIleIleThrSerThrGlnSerAlaGluAsp 1947

4225 AAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCGCAGGATTTCGGCAA 4274
     |||   :::::::|||   :::||||||:::|||:::::::::      ||
1948 LysGlyLysAsnArgPheGlnThrAlaThrLeuThrHisSerAspIleLy 1964

4275 AACCCGCAGTGCG.................GAATGGGGCGTAAACGCCG 4306
     |:::::::|||                          |||:::::|||:
1964 sAsnHisSerGlnTyrLysGlyGluSerPheGlyLeuGlyAlaSerAlaS 1981

4307 AAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGGCCG 4356
     ::|||:::|||   ||||||:::         :::|||   :::   |||
1981 erIleSerGlyLysThrLeuGly......GlnGlyAlaGlnAsnLysPro 1995

4357 CAA.................TTGGAAGCGCAGCACAGCGCGGGCATCAA 4388
     |||                 :::      ::::::||||||:::   ::
1996 GlnAsnLysHisLeuThrSerValAlaAspLysAsnSerAlaSerSerSe 2012

4389 ATTAGGCTAC 4398
     ::::||||||
2012 rValGlyTyr 2015
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01837

seq_documentation_block:

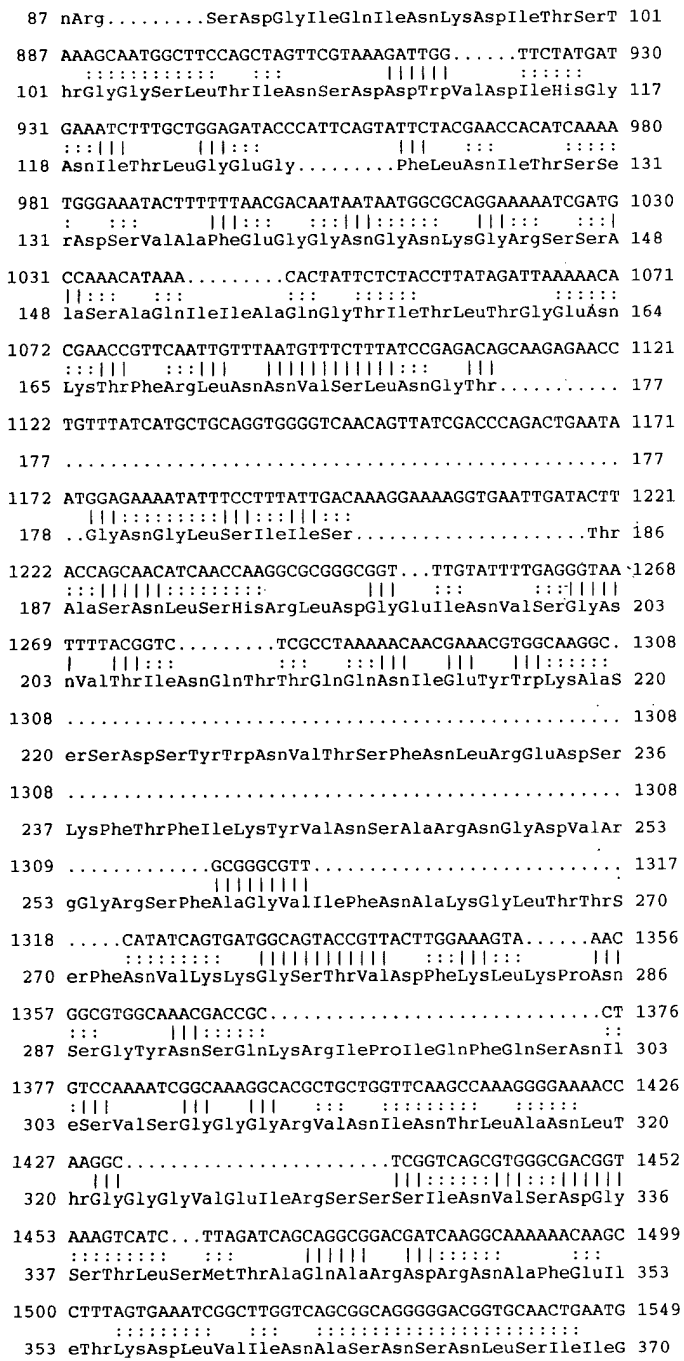
```
ID   AAB01837 standard; Protein; 1073 AA.
XX
AC   AAB01837;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain LCDC2 mature HMW2A protein, SEQ ID NO:49.
XX
KW   Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain LCDC2.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52186.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 8; Fig 23A-P; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   1073 AA;
```

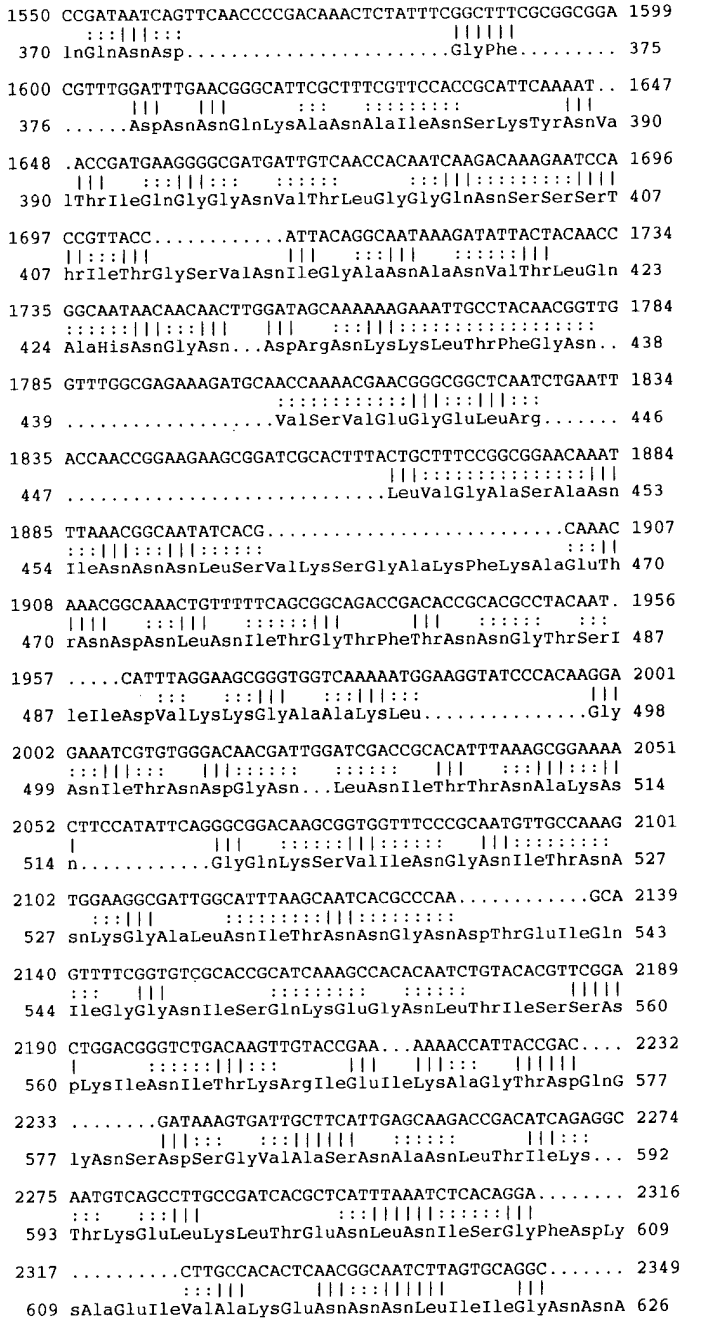
alignment_scores:

```
            Quality:  269.50              Length:   1095
              Ratio:  0.496                 Gaps:     54
 Percent Similarity:  49.589   Percent Identity: 20.639
```

alignment_block:

```
 US-09-303-518D-653 x AAB01837   ..

 Align seg 1/1  to: AAB01837  from: 1  to: 1073

     787 TCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCA 836
         ::::::||||||||||||      :::   :::|||      :::|||::
      72 AsnIleGlyAspSerGlyHis...LeuThrLeuTyrLysLysArgLysAs 87

     837 AAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAA 886
         ::::         :::      :::|||   ::::::         :::
```


87 nArg.....SerAspGlyIleGlnIleAsnLysAspIleThrSerT 101
887 AAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGG.....TTCTATGAT 930
101 hrGlyGlySerLeuThrIleAsnSerAspAspTrpValAspIleHisGly 117
931 GAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAA 980
118 AsnIleThrLeuGlyGluGly.....PheLeuAsnIleThrSerSe 131
981 TGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATG 1030
131 rAspSerValAlaPheGluGlyGlyAsnGlyAsnLysGlyArgSerSerA 148
1031 CCAAACATAAA.....CACTATTCTCTACCTTATAGATTAAAAACA 1071
148 laSerAlaGlnIleIleAlaGlnGlyThrIleThrLeuThrGlyGluAsn 164
1072 CGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACC 1121
165 LysThrPheArgLeuAsnAsnValSerLeuAsnGlyThr..... 177
1122 TGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATA 1171
177 177
1172 ATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTT 1221
178 ..GlyAsnGlyLeuSerIleIleSer.....Thr 186
1222 ACCAGCAACATCAACCAAGGCGCGGGCGGT...TTGTATTTTGAGGGTAA 1268
187 AlaSerAsnLeuSerHisArgLeuAspGlyGluIleAsnValSerGlyAs 203
1269 TTTTACGGTC.....TCGCCTAAAAACAACGAAACGTGGCAAGGC. 1308
203 nValThrIleAsnGlnThrThrGlnGlnAsnIleGluTyrTrpLysAlaS 220
1308 1308
220 erSerAspSerTyrTrpAsnValThrSerPheAsnLeuArgGluAspSer 236
1308 1308
237 LysPheThrPheIleLysTyrValAsnSerAlaArgAsnGlyAspValAr 253
1309 .....GCGGGCGTT..... 1317
253 gGlyArgSerPheAlaGlyValIlePheAsnAlaLysGlyLeuThrThrS 270
1318 .....CATATCAGTGATGGCAGTACCGTTACTTGGAAAGTA.....AAC 1356
270 erPheAsnValLysLysGlySerThrValAspPheLysLeuLysProAsn 286
1357 GGCGTGGCAAACGACCGC.....CT 1376
287 SerGlyTyrAsnSerGlnLysArgIleProIleGlnPheGlnSerAsnIl 303
1377 GTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACC 1426
303 eSerValSerGlyGlyGlyArgValAsnIleAsnThrLeuAlaAsnLeuT 320
1427 AAGGC.....TCGGTCAGCGTGGGCGACGGT 1452
320 hrGlyGlyGlyValGluIleArgSerSerSerIleAsnValSerAspGly 336
1453 AAAGTCATC...TTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGC 1499
337 SerThrLeuSerMetThrAlaGlnAlaArgAspArgAsnAlaPheGluIl 353
1500 CTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATG 1549
353 eThrLysAspLeuValIleAsnAlaSerAsnSerAsnLeuSerIleIleG 370


1550 CCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGA 1599
370 lnGlnAsnAsp.....GlyPhe..... 375
1600 CGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAAT.. 1647
376 .....AspAsnAsnGlnLysAlaAsnAlaIleAsnSerLysTyrAsnVa 390
1648 .ACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCA 1696
390 lThrIleGlnGlyGlyAsnValThrLeuGlyGlyGlnAsnSerSerSerT 407
1697 CCGTTACC.....ATTACAGGCAATAAAGATATTACTACAACC 1734
407 hrIleThrGlySerValAsnIleGlyAlaAsnAlaAsnValThrLeuGln 423
1735 GGCAATAACAACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTG 1784
424 AlaHisAsnGlyAsn...AspArgAsnLysLysLeuThrPheGlyAsn.. 438
1785 GTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAATT 1834
439 .....ValSerValGluGlyGluLeuArg..... 446
1835 ACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAAT 1884
447 .....LeuValGlyAlaSerAlaAsn 453
1885 TTAAACGGCAATATCACG.....CAAAC 1907
454 IleAsnAsnAsnLeuSerValLysSerGlyAlaLysPheLysAlaGluTh 470
1908 AAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAAT. 1956
470 rAsnAspAsnLeuAsnIleThrGlyThrPheThrAsnAsnGlyThrSerI 487
1957 .....CATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGA 2001
487 leIleAspValLysLysGlyAlaAlaLysLeu.....Gly 498
2002 GAAATCGTGTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGAAAA 2051
499 AsnIleThrAsnAspGlyAsn...LeuAsnIleThrThrAsnAlaLysAs 514
2052 CTTCCATATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAG 2101
514 n.....GlyGlnLysSerValIleAsnGlyAsnIleThrAsnA 527
2102 TGGAAGGCGATTGGCATTTAAGCAATCACGCCCAA.....GCA 2139
527 snLysGlyAlaLeuAsnIleThrAsnAsnGlyAsnAspThrGluIleGln 543
2140 GTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGA 2189
544 IleGlyGlyAsnIleSerGlnLysGluGlyAsnLeuThrIleSerSerAs 560
2190 CTGGACGGGTCTGACAAGTTGTACCGAA...AAAACCATTACCGAC.... 2232
560 pLysIleAsnIleThrLysArgIleGluIleLysAlaGlyThrAspGlnG 577
2233 .....GATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGC 2274
577 lyAsnSerAspSerGlyValAlaSerAsnAlaAsnLeuThrIleLys... 592
2275 AATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGA..... 2316
593 ThrLysGluLeuLysLeuThrGluAsnLeuAsnIleSerGlyPheAspLy 609
2317 .....CTTGCCACACTCAACGGCAATCTTAGTGCAGGC..... 2349
609 sAlaGluIleValAlaLysGluAsnAsnAsnLeuIleIleGlyAsnAsnA 626

```
2350 ..GGAGAC......ACGCACTATACGGTTACGCGCAACGCCACCCAAAAC 2391
      ||||||      :::      |||||||||   |||   :::::::::
 626 snGlyAspAsnAlaAsnAlaLysThrValThrPheAsnAsnValLysAsp 642

2392 GGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCAC 2441
     ::::::::::|||      |||                 ::::::::::||
 643 SerLysIleSerAlaAsnGly.................HisAsnValTh 653

2442 ATTAAACGGCAACACATCGGCTTCGGAC.........AATGCTTCATTTA 2482
     |||||||:::::::::::::::|||||          |||::::::   |
 653 rLeuAsnSerLysValGluThrSerAspGlyAsnSerAsnThrGluGlyA 670

2483 ATCTAAGCAACAACGCC............................GTA 2502
     ||   :::||||||||||                           |||
 670 snSerAspAsnAsnAlaGlyLeuThrIleAspAlaLysAsnValThrVal 686

2503 CAAAACGGC.....................AGTCTGACGCTTTCCGAC.. 2529
     :::|||                      ::::::|||   |||:::
 687 AsnAsnAspIleThrSerHisLysThrValAsnIleThrAlaSerGluAr 703

2530 ....AACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCT 2575
         ::::::||||||:::::::::   :::|||   :::||||||||||:
  703 gIleAspThrLysAlaAspThrThrIleAsnAlaThrThrGlyAsnValL 720

2576 CCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGA 2625
     ::|||                              :::      |||:::
 720 ysLeu............................ThrAlaValThrSer 726

2626 AAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATG 2675
        |||:::||||||      :::      :::      |||   :::
 727 AspIleGlnGlyGlyIleLysSer......AsnSerGlyAspValAsnIl 741

2676 GACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCA 2725
      |||      :::|||:::      |||:::::::::      :::::::
 741 eThrThrSerThrGlySerIleAsnGlyLysIleGluSerLysSerGlyS 758

2726 CCATTACACTCAATTCCGCCTATCGACACGATGCGGCAGGCGCGCAAACC 2775
     :::::||||||::::::::::   :::      ::::::|||      :::
 758 erValThrLeuThrAlaThrGluLysThrLeuThrValGlyAsnValSer 774

2776 GGCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATC 2825
     |||::::::                                    :::::
 775 GlyAsnThr................................ValTh 779

2826 CGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAA 2875
     :||||||         :::|||   :::   :::         |||:::|
 779 rValThrAlaAsnArgGlyAlaLeuThrThrLeuAlaGlySerThrIleA 796

2876 ACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTC 2925
     |||||      ||||||     |||           :::|||:::
 796 snGly...ThrAsnGlyValThrThrSerSerGlnSerGlyGluIleGly 811

2926 GGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTA 2975
     |||      :::|||||||         :::   ::::::   |||:::
 812 GlyGluValThrGlyLysThrValSerValThrAlaThrAlaGlySerLe 828

2976 CACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCTCGAGCAAT 3025
      |||:::      :::         |||         :::        |
 828 uThrValLysGlyGlyAlaLysIleAsnAlaThrGluGlyThrAlaThrL 845

3026 TGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAAT 3075
     |||||:::   :::||||||   :::|||      |||:::|||::::::
 845 euThrAlaSerSerGlyLysLeuThrThrGluAlaSerSerAsnIleThr 861

3076 TTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCAGCT 3125
      |||                          |||   |||   ::::::|
 862 .Ser.........................AlaLysGlyGlnValAspL 869

3126 TATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAG 3175
     ||||||||:::   :::|||   |||      |||   :::   |||
 869 euSerAlaGlnAspGlySerIleAlaGlyGlnIleSerAlaAlaAsnVal 885

3176 AGCTTTCCGACAAACTCGGC.........................AA 3197
```

```
     ::::::   |||::::::                                ::
 886 ThrLeuAsnThrThrGlyThrLeuThrThrValGluGlySerSerIleAs 902

3198 GGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCG 3247
     :|||:::|||          :::::::::|||:::::|||:::|||   :
  902 nAlaAsnGluGlyThrLeuValIleAsnAlaAsnAspAlaLysLeuAspG 919

3248 CCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATT 3297
     ::|||   ::::::::::::   ::::::::::::   :::|||
  919 lyLysAlaSerGlyAsnArgThrGluValAsnAlaThrAsnAla..... 933

3298 GCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGT.............. 3333
     :::::::::   :::::::|||   |||:::::::|||
  934 SerGlySerGlySerValThrAlaLysThrSerSerSerValAsnIleTh 950

3334 ................................GTTGCCGAACCGGCCC 3349
                                     ::::::|||   :::|
  950 rGlyAspLeuAsnThrIleAsnGlyLeuAsnIleIleSerGluAsnGlyA 967

3350 GGCAGGCA............GGCGGGGAA..................AAT 3369
     ||:::::::           |||   |||                  :::
  967 rgAsnThrValArgLeuArgGlyLysGluIleGluValLysTyrIleGln 983

3370 GCCGGCATTATGCAGGCGGAGGAA......GAGAAAAAACGGGTG...CA 3410
        |||:::   ::::::||||||      |||   |||||||||   ::
  984 ProGlyValAlaSerValGluGluValIleGluAlaLysArgValLeuGl 1000

3411 GGCGGATAAAGACACCGCC.................TTGGCGAAA.... 3438
     :      ||||||   :::                  |||||||||
 1000 uLysValLysAspLeuSerAspGluGluArgGluThrLeuAlaLysLeuG 1017

3438 ..................................................... 3438

 1017 lyValSerAlaValArgPheIleGluProAsnAsnThrIleThrValAsn 1033

3439 ...CAGCGCGAAGCGGAAACCCGGCCGGCTACC 3468
        |||:::|||      |||||||||:::::
 1034 ThrGlnAsnGluPheThrThrArgProSerSer 1044
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01836

```
seq_documentation_block:
ID   AAB01836 standard; Protein; 1079 AA.
XX
AC   AAB01836;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain LCDC2 HMW2A protein, SEQ ID NO:47.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain LCDC2.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52185.
```

```
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 23A-P; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or,
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1079 AA;


alignment_scores:
           Quality:  269.50              Length:   1095
             Ratio:  0.496                 Gaps:     54
 Percent Similarity:  49.589   Percent Identity: 20.639

alignment_block:
 US-09-303-518D-653 x AAB01836   ..

 Align seg 1/1  to: AAB01836  from: 1  to: 1079

     787 TCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCA 836
         ::::::||||||||||||      :::   :::|||      :::|||::
      78 AsnIleGlyAspSerGlyHis...LeuThrLeuTyrLysLysArgLysAs 93

     837 AAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAA 886
         ::::        :::      :::|||   ::::::          :::
      93 nArg.........SerAspGlyIleGlnIleAsnLysAspIleThrSerT 107

     887 AAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGG......TTCTATGAT 930
           ::::::::::::   :::         ||||||      ::::::
     107 hrGlyGlySerLeuThrIleAsnSerAspAspTrpValAspIleHisGly 123

     931 GAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAA 980
         :::|||      |||:::             |||   :::      :::::
     124 AsnIleThrLeuGlyGluGly.........PheLeuAsnIleThrSerSe 137

     981 TGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATG 1030
         :   :::      |||:::   :::|||::::::   |||:::   :::|
     137 rAspSerValAlaPheGluGlyGlyAsnGlyAsnLysGlyArgSerSerA 154

    1031 CCAAACATAAA.........CACTATTCTCTACCTTATAGATTAAAAACA 1071
         ||:::   :::         :::   ::::::            ::::::
     154 laSerAlaGlnIleIleAlaGlnGlyThrIleThrLeuThrGlyGluAsn 170

    1072 CGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACC 1121
         :::|||   :::|||   ||||||||||||:::   |||

     171 LysThrPheArgLeuAsnAsnValSerLeuAsnGlyThr..... 183

    1122 TGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATA 1171

     183 ..... 183

    1172 ATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTT 1221
           |||:::::::::|||:::|||:::
     184 ..GlyAsnGlyLeuSerIleIleSer.....Thr 192

    1222 ACCAGCAACATCAACCAAGGCGCGGGCGGT...TTGTATTTTGAGGGTAA 1268
         :::||||||:::::::::         |||   :::      :::||||||
     193 AlaSerAsnLeuSerHisArgLeuAspGlyGluIleAsnValSerGlyAs 209

    1269 TTTTACGGTC.........TCGCCTAAAAACAACGAAACGTGGCAAGGC. 1308
         |   |||:::          :::   :::|||   |||   |||::::::
     209 nValThrIleAsnGlnThrThrGlnGlnAsnIleGluTyrTrpLysAlaS 226

    1308 ..... 1308

     226 erSerAspSerTyrTrpAsnValThrSerPheAsnLeuArgGluAspSer 242

    1308 ..... 1308

     243 LysPheThrPheIleLysTyrValAsnSerAlaArgAsnGlyAspValAr 259

    1309 .............GCGGGCGTT..... 1317
                       |||||||||
     259 gGlyArgSerPheAlaGlyValIlePheAsnAlaLysGlyLeuThrThrS 276

    1318 .....CATATCAGTGATGGCAGTACCGTTACTTGGAAAGTA......AAC 1356
               :::::::::   ||||||||||||   :::|||:::      |||
     276 erPheAsnValLysLysGlySerThrValAspPheLysLeuLysProAsn 292

    1357 GGCGTGGCAAACGACCGC.............................CT 1376
         :::      |||::::::                            ::
     293 SerGlyTyrAsnSerGlnLysArgIleProIleGlnPheGlnSerAsnIl 309

    1377 GTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACC 1426
         :|||      |||   |||   :::   :::::::::   ::::::
     309 eSerValSerGlyGlyGlyArgValAsnIleAsnThrLeuAlaAsnLeuT 326

    1427 AAGGC.....................TCGGTCAGCGTGGGCGACGGT 1452
           |||                      |||::::::|||:::||||||
     326 hrGlyGlyGlyValGluIleArgSerSerSerIleAsnValSerAspGly 342

    1453 AAAGTCATC...TTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGC 1499
         :::::::::   :::      ||||||   |||::::::      :::
     343 SerThrLeuSerMetThrAlaGlnAlaArgAspArgAsnAlaPheGluIl 359

    1500 CTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATG 1549
         :::::::::   :::   :::::::::::::::::::::::::
     359 eThrLysAspLeuValIleAsnAlaSerAsnSerAsnLeuSerIleIleG 376

    1550 CCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGA 1599
           :::|||:::                          ||||||
     376 lnGlnAsnAsp.....................GlyPhe..... 381

    1600 CGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAAT.. 1647
           |||   |||      :::   :::::::::          |||
     382 ......AspAsnAsnGlnLysAlaAsnAlaIleAsnSerLysTyrAsnVa 396

    1648 .ACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCA 1696
          |||   :::|||:::   ::::::      :::|||:::::::::||||
     396 lThrIleGlnGlyGlyAsnValThrLeuGlyGlyGlnAsnSerSerSerT 413

    1697 CCGTTACC............ATTACAGGCAATAAAGATATTACTACAACC 1734
         ||:::|||             |||   :::|||   :::::::|||
     413 hrIleThrGlySerValAsnIleGlyAlaAsnAlaAsnValThrLeuGln 429

    1735 GGCAATAACAACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTG 1784
         :::::::|||:::|||   |||   :::|||::::::::::::::::::
     430 AlaHisAsnGlyAsn...AspArgAsnLysLysLeuThrPheGlyAsn.. 444
```

```
1785 GTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAATT 1834
                    ::::::::::::|||:::|||:::
 445 ..................ValSerValGluGlyGluLeuArg....... 452

1835 ACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAAT 1884
                              |||:::::::::::::::|||
 453 ............................LeuValGlyAlaSerAlaAsn 459

1885 TTAAACGGCAATATCACG............................CAAAC 1907
     :::|||:::|||::::::                           :::||
 460 IleAsnAsnAsnLeuSerValLysSerGlyAlaLysPheLysAlaGluTh 476

1908 AAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAAT. 1956
     ||||   :::|||   :::::::|||      |||   ::::::   :::
 476 rAsnAspAsnLeuAsnIleThrGlyThrPheThrAsnAsnGlyThrSerI 493

1957 .....CATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGA 2001
           :::   :::|||   :::|||:::               |||
 493 leIleAspValLysLysGlyAlaAlaLysLeu..............Gly 504

2002 GAAATCGTGTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGAAAA 2051
     :::|||:::   |||::::::   ::::::   |||   :::|||:::||
 505 AsnIleThrAsnAspGlyAsn...LeuAsnIleThrThrAsnAlaLysAs 520

2052 CTTCCATATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAG 2101
     |            |||   ::::::|||::::::   |||:::::::::
 520 n............GlyGlnLysSerValIleAsnGlyAsnIleThrAsnA 533

2102 TGGAAGGCGATTGGCATTTAAGCAATCACGCCCAA............GCA 2139
       :::|||      :::::::::|||:::::::::
 533 snLysGlyAlaLeuAsnIleThrAsnAsnGlyAsnAspThrGluIleGln 549

2140 GTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGA 2189
     :::   |||          :::::::::   ::::::         |||||
 550 IleGlyGlyAsnIleSerGlnLysGluGlyAsnLeuThrIleSerSerAs 566

2190 CTGGACGGGTCTGACAAGTTGTACCGAA...AAAACCATTACCGAC.... 2232
     |      ::::::|||:::      |||   |||:::   ||||||
 566 pLysIleAsnIleThrLysArgIleGluIleLysAlaGlyThrAspGlnG 583

2233 ........GATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGC 2274
             |||:::   :::||||||   ::::::      |||:::
 583 lyAsnSerAspSerGlyValAlaSerAsnAlaAsnLeuThrIleLys... 598

2275 AATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGA........ 2316
     :::   :::|||            :::||||||::::::|||
 599 ThrLysGluLeuLysLeuThrGluAsnLeuAsnIleSerGlyPheAspLy 615

2317 ..........CTTGCCACACTCAACGGCAATCTTAGTGCAGGC....... 2349
               :::|||      |||:::||||||      |||
 615 sAlaGluIleValAlaLysGluAsnAsnAsnLeuIleIleGlyAsnAsnA 632

2350 ..GGAGAC......ACGCACTATACGGTTACGCGCAACGCCACCCAAAAC 2391
       ||||||      :::      |||||||||   |||   :::::::::
 632 snGlyAspAsnAlaAsnAlaLysThrValThrPheAsnAsnValLysAsp 648

2392 GGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCAC 2441
     :::::::::|||      |||                  ::::::::::||
 649 SerLysIleSerAlaAsnGly..................HisAsnValTh 659

2442 ATTAAACGGCAACACATCGGCTTCGGAC.........AATGCTTCATTTA 2482
     |||||||:::::::::::::::::||||||         |||::::::   |
 659 rLeuAsnSerLysValGluThrSerAspGlyAsnSerAsnThrGluGlyA 676

2483 ATCTAAGCAACAACGCC.............................GTA 2502
     ||   :::|||||||||                             |||
 676 snSerAspAsnAsnAlaGlyLeuThrIleAspAlaLysAsnValThrVal 692

2503 CAAAACGGC.....................AGTCTGACGCTTTCCGAC.. 2529
     :::|||                       ::::::|||   |||:::
 693 AsnAsnAspIleThrSerHisLysThrValAsnIleThrAlaSerGluAr 709

2530 ....AACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCT 2575
         ::::::|||||::::::::::   :::|||   :::||||||||||:
  709 gIleAspThrLysAlaAspThrThrIleAsnAlaThrThrGlyAsnValL 726

2576 CCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGA 2625
     ::|||                              :::      |||:::
 726 ysLeu..............................ThrAlaValThrSer 732

2626 AAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATG 2675
        |||:::||||||      :::      :::      |||   :::
 733 AspIleGlnGlyGlyIleLysSer......AsnSerGlyAspValAsnIl 747

2676 GACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCA 2725
      |||      :::|||:::      |||:::::::::      :::::::
 747 eThrThrSerThrGlySerIleAsnGlyLysIleGluSerLysSerGlyS 764

2726 CCATTACACTCAATTCCGCCTATCGACACGATGCGGCAGGCGCGCAAACC 2775
     :::::||||||:::::::::   :::      ::::::|||      :::
 764 erValThrLeuThrAlaThrGluLysThrLeuThrValGlyAsnValSer 780

2776 GGCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATC 2825
     |||::::::                                  :::::
 781 GlyAsnThr.....................................ValTh 785

2826 CGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAA 2875
     :||||||        :::|||   :::   :::          |||:::|
 785 rValThrAlaAsnArgGlyAlaLeuThrThrLeuAlaGlySerThrIleA 802

2876 ACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTC 2925
     |||||      ||||||      |||            :::|||:::
 802 snGly...ThrAsnGlyValThrThrSerSerGlnSerGlyGluIleGly 817

2926 GGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTA 2975
     |||      :::|||||||         :::   ::::::   |||:::
 818 GlyGluValThrGlyLysThrValSerValThrAlaThrAlaGlySerLe 834

2976 CACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCTCGAGCAAT 3025
      |||:::      :::         |||         :::         |
 834 uThrValLysGlyGlyAlaLysIleAsnAlaThrGluGlyThrAlaThrL 851

3026 TGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAAT 3075
     ||||||:::   :::||||||   :::|||      |||:::|||::::::
 851 euThrAlaSerSerGlyLysLeuThrThrGluAlaSerSerAsnIleThr 867

3076 TTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCAGCT 3125
      |||                          |||   |||   :::::::|
 868 .Ser.......................AlaLysGlyGlnValAspL 875

3126 TATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAG 3175
     ||||||||:::   :::|||   |||      |||   :::   |||
 875 euSerAlaGlnAspGlySerIleAlaGlyGlnIleSerAlaAlaAsnVal 891

3176 AGCTTTCCGACAAACTCGGC..........................AA 3197
     ::::::   |||:::::::                          ::
 892 ThrLeuAsnThrThrGlyThrLeuThrThrValGluGlySerSerIleAs 908

3198 GGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCG 3247
     :|||:::|||        :::::::::|||::::::|||:::|||   :
 908 nAlaAsnGluGlyThrLeuValIleAsnAlaAsnAspAlaLysLeuAspG 925

3248 CCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATT 3297
     ::|||   ::::::::::::   ::::::::::::   :::|||
 925 lyLysAlaSerGlyAsnArgThrGluValAsnAlaThrAsnAla..... 939

3298 GCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGT.............. 3333
     :::::::::   ::::::|||   |||:::::::|||
 940 SerGlySerGlySerValThrAlaLysThrSerSerSerValAsnIleTh 956

3334 ..............................GTTGCCGAACCGGCCC 3349
                                   :::::::|||   :::|
 956 rGlyAspLeuAsnThrIleAsnGlyLeuAsnIleIleSerGluAsnGlyA 973

3350 GGCAGGCA............GGCGGGGAA.................AAT 3369
```

```
     ||::::::           |||   |||                  :::
 973 rgAsnThrValArgLeuArgGlyLysGluIleGluValLysTyrIleGln 989

3370 GCCGGCATTATGCAGGCGGAGGAA......GAGAAAAAACGGGTG...CA 3410
        |||:::   ::::::||||||      |||   |||||||||   ::
 990 ProGlyValAlaSerValGluGluValIleGluAlaLysArgValLeuGl 1006

3411 GGCGGATAAAGACACCGCC.................TTGGCGAAA.... 3438
     :      ||||||   :::                  |||||||||
1006 uLysValLysAspLeuSerAspGluGluArgGluThrLeuAlaLysLeuG 1023

3438 ..................................................... 3438

1023 lyValSerAlaValArgPheIleGluProAsnAsnThrIleThrValAsn 1039

3439 ...CAGCGCGAAGCGGAAACCCGGCCGGCTACC 3468
        |||:::|||      |||||||||:::::
1040 ThrGlnAsnGluPheThrThrArgProSerSer 1050
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR63505

seq_documentation_block:

```
ID   AAR63505 standard; Protein; 1536 AA.
XX
AC   AAR63505;
XX
DT   25-JUN-1995  (first entry)
XX
DE   Haemophilus high molecular weight protein HMW1.
XX
KW   High molecular weight protein; HMW1; protective vaccine; otitis;
KW   sinusitis; bronchitis; Hib.
XX
OS   Haemophilus.
XX
PN   WO9421290-A.
XX
PD   29-SEP-1994.
XX
PF   15-MAR-1994;   94WO-US02550.
XX
PR   16-MAR-1993;   93US-0038682.
XX
PA   (BARE/) BARENKAMP S J.
PA   (SGEM/) ST GEME J W.
XX
PI   Barenkamp SJ,  St GEME JW;
XX
DR   WPI; 1994-316665/39.
DR   Q-PSDB; Q72293.
XX
PT   New immunogenic high mol. wt. proteins of non typeable
PT   Haemophilus - useful in protective vaccines
XX
PS    Claim 2; Page 31; 127pp; English.
XX
CC   The HMW1 protein encoded by this sequence is useful in a vaccine to
CC   protect against disease caused by non-typeable Haemophilus which are
CC   not controlled by H. influenzae type b (Hib) vaccines.  The encoded
CC   protein can also be used as a carrier for protective Hib
CC   polysaccharide (in a conjugate vaccin against meningitis) or for
CC   other antigens, haptens, etc.
XX
SQ   Sequence   1536 AA;
```

alignment_scores:

```
            Quality:  268.50            Length:   1217
              Ratio:  0.465               Gaps:     60
 Percent Similarity:  47.494   Percent Identity: 20.624
```

alignment_block:

```
 US-09-303-518D-653 x AAR63505   ..
```

```
Align seg 1/1  to: AAR63505  from: 1  to: 1536

      1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCTAA 50
        |||   |||   ||||||   |||   :::|||
    327 MetIleThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLe 343

     51 AACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGT 100
         :::|||:::
    343 uSerGlyLys..................................... 346

    101 TCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATCAAC 150
                              :::::::|||:::|||||||:::|||
    347 ....................GluGlyGlyGluThrTyrLeuGly..... 354

    151 TACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGG 200
                              ::::::::::::|||:::         ||
    355 ....................GlyAspGluArgGlyGlu.........Gl 361

    201 GGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAAT 250
        |   |||   |||::::::   :::|||                     |
    361 yLysLysGlyIleGlnLeuAlaLysLys.................ThrS 372

    251 CGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAACGGC 300
        ||:::   |||:::       |||:::   |||      :::::::::|||
    372 erLeuGluLysGlySerThrIleAsnValSerGlyLysGluLysGlyGly 388

    301 GTGGCGGCATTGGCGGGCGATCAATATATTGTG................. 333
           |||   :::   ||||||      ::::::
    389 ArgAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAl 405

    334 .............AGCGTGGCACATAACGGCGGCTAT.............. 357
                     ::::::|||   :::||||||:::
    405 aGlnGlySerGlyAspIleAlaLysThrGlyGlyPheValGluThrSerG 422

    357 ..................................................... 357

    422 lyHisAspLeuPheIleLysAspAsnAlaIleValAspAlaLysGluTrp 438

    358 ............AACAATGTTGATTTTGGTGCGGAG.............. 381
                    :::||||||:::::::::||||||
    439 LeuLeuAspProAspAsnValSerIleAsnAlaGluThrAlaGlyArgSe 455

    382 ....GGAAGCAATCCCGATCAGCACCGC.........TTTTCTTACCAAA 418
              |||:::   |||::::::                |||   :::
    455 rAsnThrSerGluAspAspGluTyrThrGlySerGlyAsnSerAlaSerT 472

    419 TTGTGAAAAGAAATAAT................................ 435
          |||||||||:::
    472 hrProLysArgAsnLysGluLysThrThrLeuThrAsnThrThrLeuGlu 488

    436 ......TATAAAGCAGGGACT........................... 450
               |||   ||||||
    489 SerIleLeuLysLysGlyThrPheValAsnIleThrAlaAsnGlnArgIl 505

    450 ..................................................... 450

    505 eTyrValAsnSerSerIleAsnLeuSerAsnGlySerLeuThrLeuTrpS 522

    451 ..AACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCACAAA 498
          :::|||:::      ||||||              ::::::::::::
    522 erGluGlyArgSerGlyGlyGly............ValGluIleAsnAsn 534

    499 TTTGTCACA.........GATGCAGAACCTGTTGAGATGACCAGTTATAT 539
           :::|||          |||::::::   :::::::::|||   |||
    535 AspIleThrThrGlyAspAspThrArgGlyAlaAsnLeuThrIleTyrSe 551

    540 GGATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAA 589
           ||||||      :::|||::::::|||         ::::::   :
    551 rGlyGlyTrp......ValAspValHisLys.........AsnIleSerL 563

    590 TCGGAGCA...GGCAGACAATATTGGCGGTCTGATGAAGACGAACCCAAT 636
        ::||||||   |||:::            :::   :::|||
```

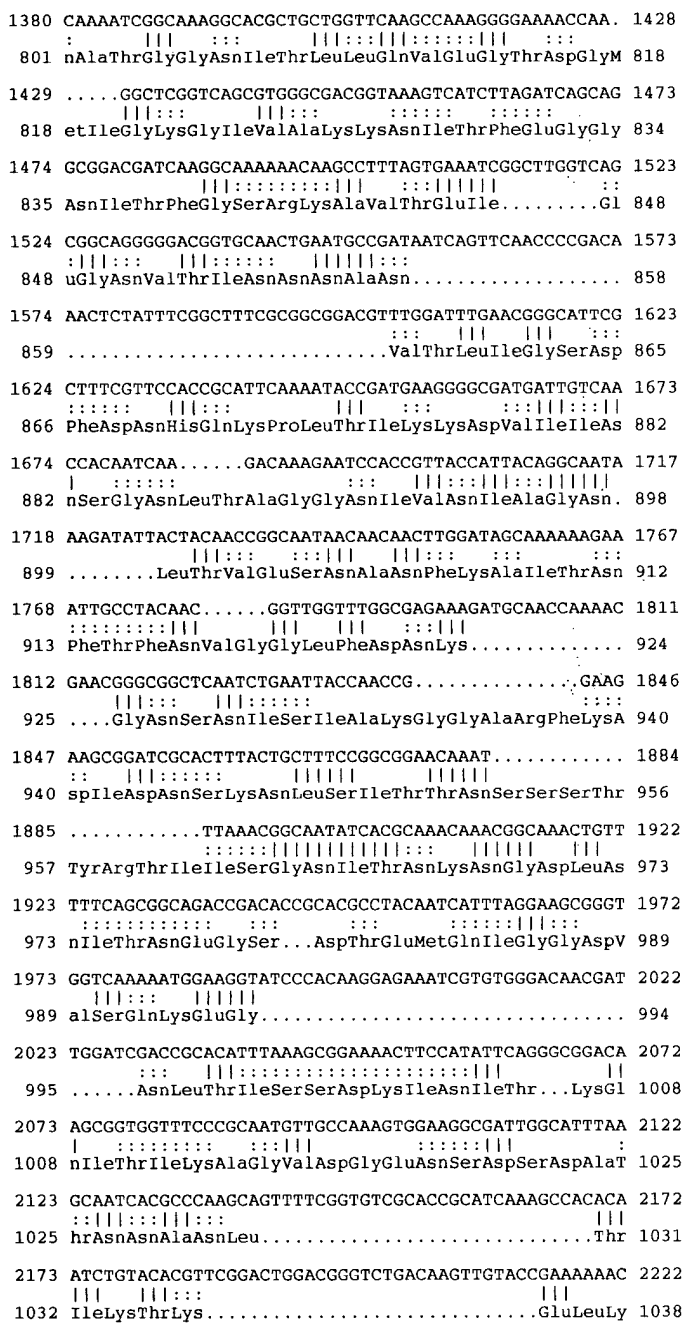
```
 563 euGlyAlaGlnGlyAsnIleAsnIleThrAlaLysGlnAspIleAlaPhe 579

 637 AACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGTCGG 686
     ::::::   |||:::      |||
 580 GluLysGlySerAsnGlnValIle..................... 587

 687 TGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAG 736
                              :::|||   ||||||::::::   |
 588 .......................ThrGlyGlnGlyThrIleThrSerG 596

 737 GTAGCGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGC 786
     ||:::::::|||                    ||||||      ::::::
 596 lyAsnGlnLys.................GlyPheArgPheAsnAsnVal 606

 787 TCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCA 836
     |||::::::   :::||||||   :::   :::          :::   ::
 607 SerLeuAsnGlyThrGlySerGlyLeuGlnPheThrThrLysArgThrAs 623

 837 AAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAA 886
     :|||:::   |||::::::   ::::::                  ||||
 623 nLysTyrAlaIleThrAsnLysPheGluGlyThrLeuAsnIleSerGlyL 640

 887 AAAGCAAT...GGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAA 933
     ||   |||   ::::::   |||   :::::::::      ||||||:::
 640 ysValAsnIleSerMetValLeuProLysAsnGluSerGlyTyrAspLys 656

 934 ATCTTTGCTGGAGATACCCAT......TCAGTATTCTACGAACCACATCA 977
        |||   |||   |||:::          :::      :::      ::
 657 ...PheLysGlyArgThrTyrTrpAsnLeuThrSerLeuAsnValSerGl 672

 978 AAATGGGAAATACTTTTTTAACGACAATAATAATGGC.........GCAG 1018
     ::::|||::::::   ::::::   :::::::::|||         ||||
 672 uSerGlyGluPheAsnLeuThrIleAspSerArgGlySerAspSerAlaG 689

1019 GAAAAATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAA 1068
     ||   :::          ::::::      ::::::   ::::::
 689 lyThrLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysAsp 705

1069 ACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGA 1118
     |||               |||||||||:::   :::      ::::::
 706 Thr............ThrPheAsnValGluArgAsnAlaArgValAsnPh 718

1119 ACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGA 1168
        :::      |||     |||:::|||:::|||         ||||
 718 eAspIle...LysAlaProIleGlyIleAsnLysTyr...SerSerLeuA 733

1169 AT............AATGGAGAAAATATTTCCTTTATTGACAAAGGAAAA 1206
     ||           ||||||   |||||||||           |||:::
 733 snTyrAlaSerPheAsnGly...AsnIleSerValSerGlyGlyGlySer 748

1207 GGTGAATTG...ATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTT 1253
        ::::::   :::|||:::|||:::   |||      :::   |||::
 749 ValAspPheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyVa 765

1254 GTATTTTGAGGGTAATTTTACGGTCTCGCCTAAAAACAACGAAACGTGGC 1303
     :   ::::::::::::::::
 765 lValIleAsnSerLysTyr............................... 771

1304 AAGGCGCGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTA 1353
                  :::::::|||   ||||||::::::   :::|||:::
 772 ...........PheAsnValSerThrGlySerSerLeuArgPheLysThr 784

1354 AACGGC.....................GTGGCAAACGAC...CGCCTGTC 1379
     :::|||                          :::   :::|||      |||::
 785 SerGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeuAs 801

1380 CAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAA. 1428
     :      |||   :::      |||:::|||::::::|||   :::
 801 nAlaThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAspGlyM 818

1429 .....GGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAG 1473
          |||:::      |||:::      ::::::   ::::::
 818 etIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGlyGly 834
```

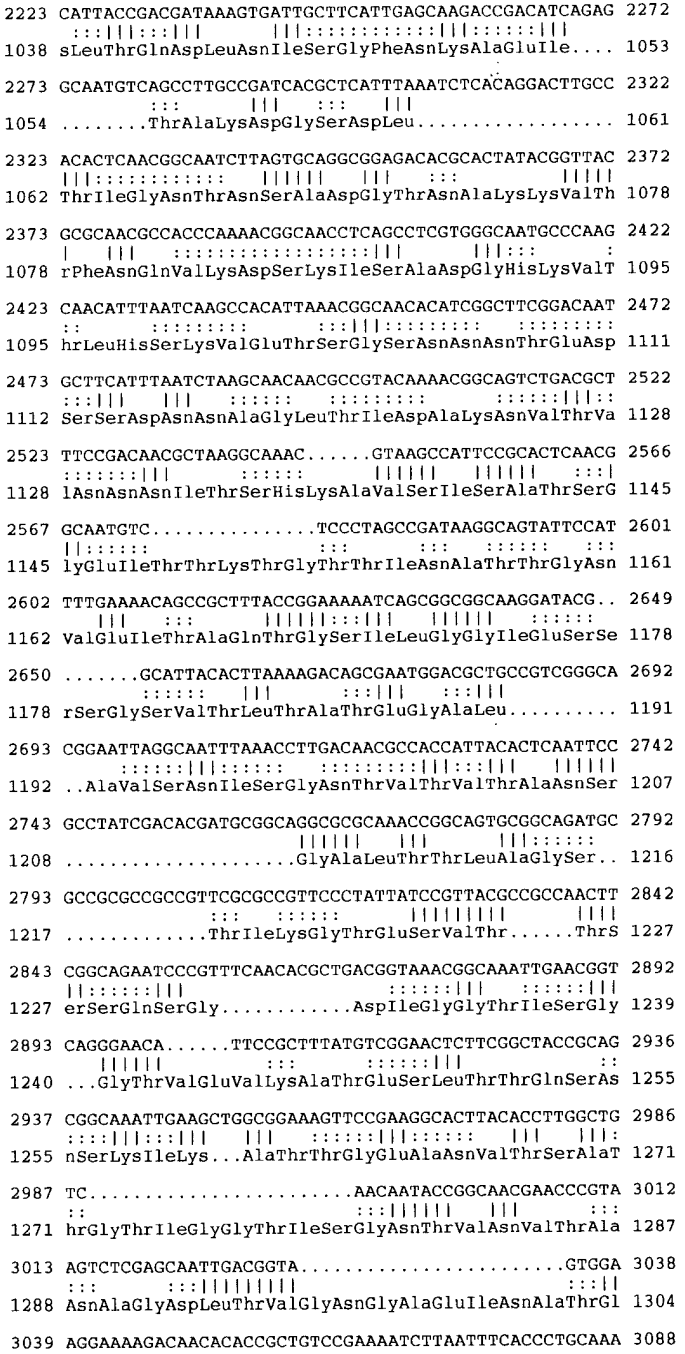
```
1474 GCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAG 1523
                  |||:::::::::|||   :::|||||||         ::
 835 AsnIleThrPheGlySerArgLysAlaValThrGluIle.........Gl 848

1524 CGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACA 1573
     :|||:::   |||::::::   ||||||:::
 848 uGlyAsnValThrIleAsnAsnAsnAlaAsn..................... 858

1574 AACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCG 1623
                                    :::   |||   |||   :::
 859 ............................ValThrLeuIleGlySerAsp 865

1624 CTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAA 1673
     ::::::   |||:::          |||   :::      :::|||:::||
 866 PheAspAsnHisGlnLysProLeuThrIleLysLysAspValIleIleAs 882

1674 CCACAATCAA......GACAAAGAATCCACCGTTACCATTACAGGCAATA 1717
     |   ::::::                :::   |||:::|||:::||||||
 882 nSerGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsn. 898

1718 AAGATATTACTACAACCGGCAATAACAACAACTTGGATAGCAAAAAAGAA 1767
                  |||:::   :::|||   |||:::   :::      :::
 899 ........LeuThrValGluSerAsnAlaAsnPheLysAlaIleThrAsn 912

1768 ATTGCCTACAAC......GGTTGGTTTGGCGAGAAAGATGCAACCAAAAC 1811
     :::::::::|||      |||   |||   :::|||
 913 PheThrPheAsnValGlyGlyLeuPheAspAsnLys.............. 924

1812 GAACGGGCGGCTCAATCTGAATTACCAACCG..............GAAG 1846
          |||:::   |||::::::                          ::::
 925 ....GlyAsnSerAsnIleSerIleAlaLysGlyGlyAlaArgPheLysA 940

1847 AAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAAT............ 1884
     ::   |||::::::      ||||||      ||||||
 940 spIleAspAsnSerLysAsnLeuSerIleThrThrAsnSerSerSerThr 956

1885 ............TTAAACGGCAATATCACGCAAACAAACGGCAAACTGTT 1922
                  ::::::||||||||||||||:::   ||||||   |||
 957 TyrArgThrIleIleSerGlyAsnIleThrAsnLysAsnGlyAspLeuAs 973

1923 TTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGT 1972
      ::::::::::::   :::      :::      :::::::|||:::
 973 nIleThrAsnGluGlySer...AspThrGluMetGlnIleGlyGlyAspV 989

1973 GGTCAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGAT 2022
       |||:::   ||||||
 989 alSerGlnLysGluGly.............................. 994

2023 TGGATCGACCGCACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACA 2072
           :::   |||:::::::::::::::::::::::|||         ||
 995 ......AsnLeuThrIleSerSerAspLysIleAsnIleThr...LysGl 1008

2073 AGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAA 2122
     |   :::::::::   :::|||          ::::::|||        :
1008 nIleThrIleLysAlaGlyValAspGlyGluAsnSerAspSerAspAlaT 1025

2123 GCAATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACA 2172
     ::|||:::|||:::                                 |||
1025 hrAsnAsnAlaAsnLeu............................Thr 1031

2173 ATCTGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAAC 2222
     |||   |||:::                                  |||
1032 IleLysThrLys............................GluLeuLy 1038

2223 CATTACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAG 2272
      :::|||:::|||      |||:::::::::::::|||:::::::|||
1038 sLeuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluIle.... 1053

2273 GCAATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCC 2322
           :::      |||   :::   |||
1054 ........ThrAlaLysAspGlySerAspLeu.................. 1061
```

```
2323 ACACTCAACGGCAATCTTAGTGCAGGCGGAGACACGCACTATACGGTTAC 2372
     |||:::::::::::::   ||||||   |||   :::          |||||
1062 ThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysLysValTh 1078

2373 GCGCAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAG 2422
     |   |||   ::::::::::::::::::|||      |||:::      :
1078 rPheAsnGlnValLysAspSerLysIleSerAlaAspGlyHisLysValT 1095

2423 CAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGACAAT 2472
     ::      :::::::::      :::|||:::::::::   :::::::::
1095 hrLeuHisSerLysValGluThrSerGlySerAsnAsnAsnThrGluAsp 1111

2473 GCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCT 2522
     :::|||   |||   ::::::   :::::::::      :::::::|||::
1112 SerSerAspAsnAsnAlaGlyLeuThrIleAspAlaLysAsnValThrVa 1128

2523 TTCCGACAACGCTAAGGCAAAC......GTAAGCCATTCCGCACTCAACG 2566
     ::::::::|||      ::::::      ||||||   ||||||   :::|
1128 lAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSerG 1145

2567 GCAATGTC..............TCCCTAGCCGATAAGGCAGTATTCCAT 2601
     ||::::::               :::      :::   ::::::   :::
1145 lyGluIleThrThrLysThrGlyThrThrIleAsnAlaThrThrGlyAsn 1161

2602 TTTGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACG.. 2649
        |||   :::      ||||||:::|||   ||||||   ::::::
1162 ValGluIleThrAlaGlnThrGlySerIleLeuGlyGlyIleGluSerSe 1178

2650 .......GCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCA 2692
             ::::::   |||      :::|||   :::|||
1178 rSerGlySerValThrLeuThrAlaThrGluGlyAlaLeu.......... 1191

2693 CGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCC 2742
          ::::::|||::::::   ::::::::::|||:::|||   ||||||
1192 ..AlaValSerAsnIleSerGlyAsnThrValThrValThrAlaAsnSer 1207

2743 GCCTATCGACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGC 2792
                           ||||||   |||      |||::::::
1208 .....................GlyAlaLeuThrThrLeuAlaGlySer.. 1216

2793 GCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTT 2842
                  :::   ::::::      |||||||||      ||||
1217 .............ThrIleLysGlyThrGluSerValThr......ThrS 1227

2843 CGGCAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGT 2892
     ||:::::::|||                  ::::::|||   :::::::|||
1227 erSerGlnSerGly............AspIleGlyGlyThrIleSerGly 1239

2893 CAGGGAACA......TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAG 2936
        ||||||        :::      ::::::|||            ::
1240 ...GlyThrValGluValLysAlaThrGluSerLeuThrThrGlnSerAs 1255

2937 CGGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTG 2986
     :::|||:::|||   |||   ::::::|||::::::   |||   |||:
1255 nSerLysIleLys...AlaThrThrGlyGluAlaAsnValThrSerAlaT 1271

2987 TC.......................AACAATACCGGCAACGAACCCGTA 3012
     ::                      :::||||||   |||      :::
1271 hrGlyThrIleGlyGlyThrIleSerGlyAsnThrValAsnValThrAla 1287

3013 AGTCTCGAGCAATTGACGGTA.......................GTGGA 3038
     :::      :::|||||||||                      :::||
1288 AsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGl 1304

3039 AGGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAA 3088
     ||||      :::         |||::::::         |||      :
1304 uGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrThrGluAlaS 1321

3089 ACGAACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGAC 3138
     :::::|||:::   ::: |||   |||   ::::::|||||||||:::
1321 erSerHisIleThrSer.AlaLysGlyGlnValAsnLeuSerAlaGlnAs 1337

3139 GGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACA 3187
```

```
        :::|||   |||      |||::::::   |||   ::::::   |||
    1337 pGlySerValAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThr 1353

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01846

seq_documentation_block:
ID   AAB01846 standard; Protein; 1536 AA.
XX
AC   AAB01846;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain 12 HMW1A protein, SEQ ID NO:67.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain 12.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52195.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Example 16; Fig 28A-Q; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1536 AA;
```

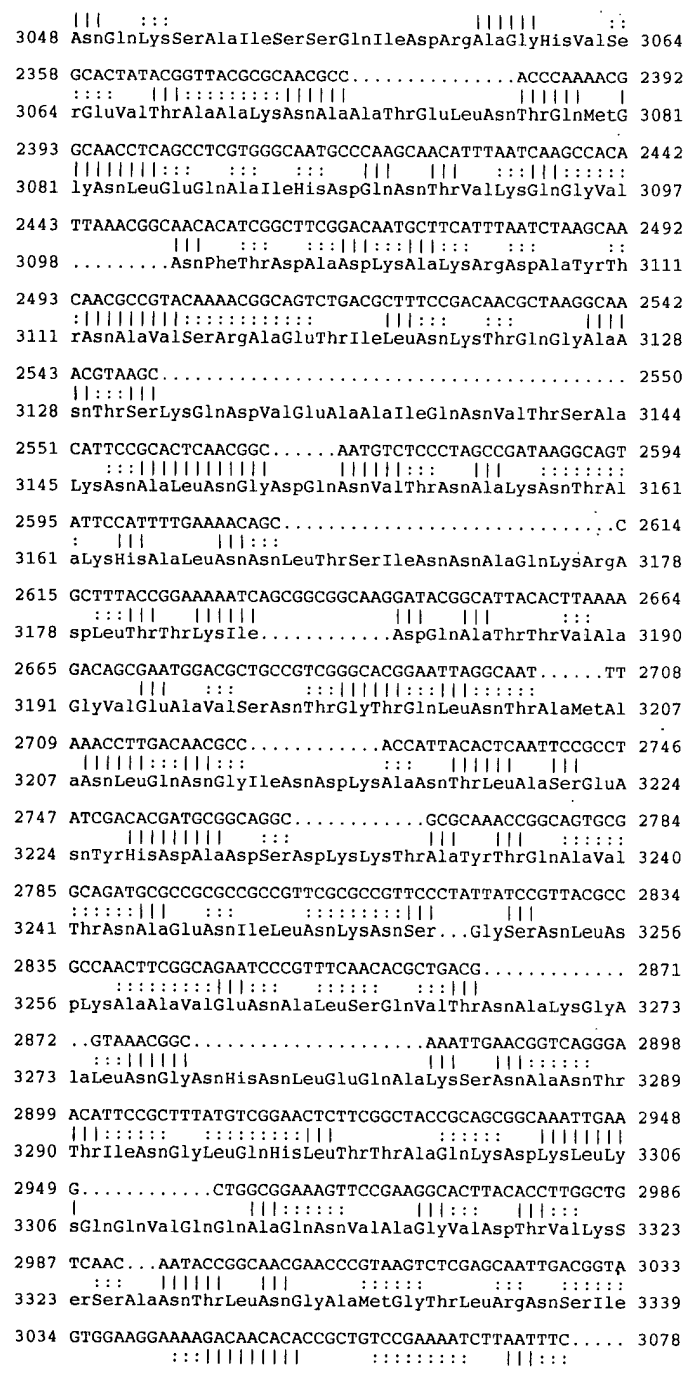
```
alignment_scores:
           Quality:  268.50              Length:   1217
             Ratio:  0.465                Gaps:     60
 Percent Similarity:  47.494   Percent Identity: 20.624

alignment_block:
 US-09-303-518D-653 x AAB01846   ..

 Align seg 1/1  to: AAB01846  from: 1  to: 1536

       1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCTAA 50
         |||   |||   ||||||   |||   :::|||
     327 MetIleThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLe 343

      51 AACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGT 100
         :::|||:::
     343 uSerGlyLys..................................... 346

     101 TCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATCAAC 150
                         ::::::|||:::||||||:::|||
     347 ...................GluGlyGlyGluThrTyrLeuGly...... 354

     151 TACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGG 200
                           :::::::::::::|||:::         ||
     355 ....................GlyAspGluArgGlyGlu.........Gl 361

     201 GGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAAT 250
         |   |||   |||::::::   :::|||                    |
     361 yLysLysGlyIleGlnLeuAlaLysLys.................ThrS 372

     251 CGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAACGGC 300
         ||:::   |||:::      |||:::   |||      :::::::::|||
     372 erLeuGluLysGlySerThrIleAsnValSerGlyLysGluLysGlyGly 388

     301 GTGGCGGCATTGGCGGGCGATCAATATATTGTG.................. 333
            |||   :::   ||||||      ::::::
     389 ArgAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAl 405

     334 .............AGCGTGGCACATAACGGCGGCTAT.............. 357
                         ::::::|||   :::|||||||:::
     405 aGlnGlySerGlyAspIleAlaLysThrGlyGlyPheValGluThrSerG 422

     357 ................................................... 357

     422 lyHisAspLeuPheIleLysAspAsnAlaIleValAspAlaLysGluTrp 438

     358 ............AACAATGTTGATTTTGGTGCGGAG.............. 381
                      :::||||||:::::::::||||||
     439 LeuLeuAspProAspAsnValSerIleAsnAlaGluThrAlaGlyArgSe 455

     382 ....GGAAGCAATCCCGATCAGCACCGC.........TTTTCTTACCAAA 418
               |||:::   |||::::::               |||   :::
     455 rAsnThrSerGluAspAspGluTyrThrGlySerGlyAsnSerAlaSerT 472

     419 TTGTGAAAAGAAATAAT................................. 435
               ||||||||||:::
     472 hrProLysArgAsnLysGluLysThrThrLeuThrAsnThrThrLeuGlu 488

     436 ......TATAAAGCAGGGACT............................ 450
                 |||   ||||||
     489 SerIleLeuLysLysGlyThrPheValAsnIleThrAlaAsnGlnArgIl 505

     450 ................................................... 450

     505 eTyrValAsnSerSerIleAsnLeuSerAsnGlySerLeuThrLeuTrpS 522

     451 ..AACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCACAAA 498
           :::|||:::      ||||||               ::::::::::::
     522 erGluGlyArgSerGlyGlyGly............ValGluIleAsnAsn 534

     499 TTTGTCACA.........GATGCAGAACCTGTTGAGATGACCAGTTATAT 539
           :::|||          |||::::::   :::::::::|||   |||

     535 AspIleThrThrGlyAspAspThrArgGlyAlaAsnLeuThrIleTyrSe 551

     540 GGATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAA 589
             ||||||      :::|||:::::::|||         ::::::   :
     551 rGlyGlyTrp......ValAspValHisLys.........AsnIleSerL 563

     590 TCGGAGCA...GGCAGACAATATTGGCGGTCTGATGAAGACGAACCCAAT 636
         ::||||||   |||:::              :::   :::|||
     563 euGlyAlaGlnGlyAsnIleAsnIleThrAlaLysGlnAspIleAlaPhe 579

     637 AACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGTCGG 686
         ::::::   |||:::      |||
     580 GluLysGlySerAsnGlnValIle........................... 587

     687 TGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAG 736
                                  :::|||   ||||||::::::   |
     588 .........................ThrGlyGlnGlyThrIleThrSerG 596

     737 GTAGCGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGC 786
         ||:::::|||                  ||||||      ::::::
     596 lyAsnGlnLys..................GlyPheArgPheAsnAsnVal 606

     787 TCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCA 836
         |||::::::   :::||||||   :::   :::          :::   ::
     607 SerLeuAsnGlyThrGlySerGlyLeuGlnPheThrThrLysArgThrAs 623

     837 AAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAA 886
         :|||:::   |||::::::   ::::::                  ||||
     623 nLysTyrAlaIleThrAsnLysPheGluGlyThrLeuAsnIleSerGlyL 640

     887 AAAGCAAT...GGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAA 933
         ||   |||   ::::::   |||   :::::::::      ||||||:::
     640 ysValAsnIleSerMetValLeuProLysAsnGluSerGlyTyrAspLys 656

     934 ATCTTTGCTGGAGATACCCAT......TCAGTATTCTACGAACCACATCA 977
            |||   |||   |||:::        :::      :::      ::
     657 ...PheLysGlyArgThrTyrTrpAsnLeuThrSerLeuAsnValSerGl 672

     978 AAATGGGAAATACTTTTTTAACGACAATAATAATGGC.........GCAG 1018
         ::::|||::::::   ::::::   :::::::::|||         ||||
     672 uSerGlyGluPheAsnLeuThrIleAspSerArgGlySerAspSerAlaG 689

    1019 GAAAAATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAA 1068
         ||   :::         ::::::      ::::::   ::::::
     689 lyThrLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysAsp 705

    1069 ACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGA 1118
         |||                |||||||||:::   :::      ::::::
     706 Thr............ThrPheAsnValGluArgAsnAlaArgValAsnPh 718

    1119 ACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGA 1168
            :::      |||      |||:::|||:::|||         ||||
     718 eAspIle...LysAlaProIleGlyIleAsnLysTyr...SerSerLeuA 733

    1169 AT............AATGGAGAAAATATTTCCTTTATTGACAAAGGAAAA 1206
         ||            ||||||   |||||||||            |||:::
     733 snTyrAlaSerPheAsnGly...AsnIleSerValSerGlyGlyGlySer 748

    1207 GGTGAATTG...ATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTT 1253
            ::::::   :::|||:::|||:::   |||      :::   |||::
     749 ValAspPheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyVa 765

    1254 GTATTTTGAGGGTAATTTTACGGTCTCGCCTAAAAACAACGAAACGTGGC 1303
         :   ::::::::::::::::
     765 lValIleAsnSerLysTyr............................... 771

    1304 AAGGCGCGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTA 1353
                         :::::::|||   ||||||::::::   :::|||:::
     772 ...........PheAsnValSerThrGlySerSerLeuArgPheLysThr 784

    1354 AACGGC.....................GTGGCAAACGAC...CGCCTGTC 1379
         :::|||                     :::   :::|||      ||||::
     785 SerGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeuAs 801
```


1380 CAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAA. 1428
801 nAlaThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAspGlyM 818
1429 .....GGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAG 1473
818 etIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGlyGly 834
1474 GCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAG 1523
835 AsnIleThrPheGlySerArgLysAlaValThrGluIle.....Gl 848
1524 CGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACA 1573
848 uGlyAsnValThrIleAsnAsnAsnAlaAsn..... 858
1574 AACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCG 1623
859 .....ValThrLeuIleGlySerAsp 865
1624 CTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAA 1673
866 PheAspAsnHisGlnLysProLeuThrIleLysLysAspValIleIleAs 882
1674 CCACAATCAA......GACAAAGAATCCACCGTTACCATTACAGGCAATA 1717
882 nSerGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsn. 898
1718 AAGATATTACTACAACCGGCAATAACAACAACTTGGATAGCAAAAAAGAA 1767
899 ....LeuThrValGluSerAsnAlaAsnPheLysAlaIleThrAsn 912
1768 ATTGCCTACAAC......GGTTGGTTTGGCGAGAAAGATGCAACCAAAAC 1811
913 PheThrPheAsnValGlyGlyLeuPheAspAsnLys..... 924
1812 GAACGGGCGGCTCAATCTGAATTACCAACCG.....GAAG 1846
925 ....GlyAsnSerAsnIleSerIleAlaLysGlyGlyAlaArgPheLysA 940
1847 AAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAAT..... 1884
940 spIleAspAsnSerLysAsnLeuSerIleThrThrAsnSerSerSerThr 956
1885 .....TTAAACGGCAATATCACGCAAACAAACGGCAAACTGTT 1922
957 TyrArgThrIleIleSerGlyAsnIleThrAsnLysAsnGlyAspLeuAs 973
1923 TTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGT 1972
973 nIleThrAsnGluGlySer...AspThrGluMetGlnIleGlyGlyAspV 989
1973 GGTCAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGAT 2022
989 alSerGlnLysGluGly..... 994
2023 TGGATCGACCGCACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACA 2072
995 ......AsnLeuThrIleSerSerAspLysIleAsnIleThr...LysGl 1008
2073 AGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAA 2122
1008 nIleThrIleLysAlaGlyValAspGlyGluAsnSerAspSerAspAlaT 1025
2123 GCAATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACA 2172
1025 hrAsnAsnAlaAsnLeu.....Thr 1031
2173 ATCTGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAAC 2222
1032 IleLysThrLys.....GluLeuLy 1038


2223 CATTACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAG 2272
1038 sLeuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluIle.... 1053
2273 GCAATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCC 2322
1054 .....ThrAlaLysAspGlySerAspLeu..... 1061
2323 ACACTCAACGGCAATCTTAGTGCAGGCGGAGACACGCACTATACGGTTAC 2372
1062 ThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysLysValTh 1078
2373 GCGCAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAG 2422
1078 rPheAsnGlnValLysAspSerLysIleSerAlaAspGlyHisLysValT 1095
2423 CAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGACAAT 2472
1095 hrLeuHisSerLysValGluThrSerGlySerAsnAsnAsnThrGluAsp 1111
2473 GCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCT 2522
1112 SerSerAspAsnAsnAlaGlyLeuThrIleAspAlaLysAsnValThrVa 1128
2523 TTCCGACAACGCTAAGGCAAAC......GTAAGCCATTCCGCACTCAACG 2566
1128 lAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSerG 1145
2567 GCAATGTC.....TCCCTAGCCGATAAGGCAGTATTCCAT 2601
1145 lyGluIleThrThrLysThrGlyThrThrIleAsnAlaThrThrGlyAsn 1161
2602 TTTGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACG.. 2649
1162 ValGluIleThrAlaGlnThrGlySerIleLeuGlyGlyIleGluSerSe 1178
2650 .....GCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCA 2692
1178 rSerGlySerValThrLeuThrAlaThrGluGlyAlaLeu..... 1191
2693 CGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCC 2742
1192 ..AlaValSerAsnIleSerGlyAsnThrValThrValThrAlaAsnSer 1207
2743 GCCTATCGACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGC 2792
1208 .....GlyAlaLeuThrThrLeuAlaGlySer.. 1216
2793 GCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTT 2842
1217 .....ThrIleLysGlyThrGluSerValThr.....ThrS 1227
2843 CGGCAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGT 2892
1227 erSerGlnSerGly.....AspIleGlyGlyThrIleSerGly 1239
2893 CAGGGAACA.....TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAG 2936
1240 ...GlyThrValGluValLysAlaThrGluSerLeuThrThrGlnSerAs 1255
2937 CGGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTG 2986
1255 nSerLysIleLys...AlaThrThrGlyGluAlaAsnValThrSerAlaT 1271
2987 TC.....AACAATACCGGCAACGAACCCGTA 3012
1271 hrGlyThrIleGlyGlyThrIleSerGlyAsnThrValAsnValThrAla 1287
3013 AGTCTCGAGCAATTGACGGTA.....GTGGA 3038
1288 AsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGl 1304
3039 AGGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAA 3088

```
      ||||      :::         |||:::::::          |||    .  :
  1304 uGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrThrGluAlaS 1321

  3089 ACGAACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGAC 3138
       :::::|||:::   ::: |||   |||   ::::::|||||||||:::
  1321 erSerHisIleThrSer.AlaLysGlyGlnValAsnLeuSerAlaGlnAs 1337

  3139 GGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACA 3187
        :::|||   |||      |||::::::   |||   ::::::   |||
  1337 pGlySerValAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThr 1353

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37017

seq_documentation_block:
ID   AAU37017 standard; Protein; 5795 AA.
XX
AC   AAU37017;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #1187.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS54876.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 12610; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   5795 AA;


alignment_scores:
           Quality:  267.50              Length:   1561
             Ratio:  0.349                 Gaps:     82
 Percent Similarity:  49.071   Percent Identity: 21.973

alignment_block:
 US-09-303-518D-653 x AAU37017   ..

 Align seg 1/1  to: AAU37017  from: 1  to: 5795

     274 GATTTTTCTGTGGTATCGCGTAACGGC..................... 300
         |||      |||::::::||||||||||
    2244 AspLysLeuValIleLysArgAsnGlyAsnValValThrIlePheThrAr 2260

     301 .........................................GTGG 304
                                                     ||||
    2260 gArgAsnAsnThrSerProTrpValLysGluAlaSerAlaAlaThrValA 2277

     305 CGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGC 354
         ||::::::||||||   :::      ::::::|||:::     |||
    2277 laGlyIleAlaGlyThrAsnAsnGlyIleThrValThr...AlaGlyThr 2292

     355 TATAACAATGTTGAT............TTTGGTGCGGAGGGAAGC..... 387
         :::|||   :::|||               :::::::::||||||
    2293 PheAsnProAlaAspThrIleGlnValValAlaThrGlnGlySerGlyGl 2309

     388 .......AATCCCGATCAGCACCGCTTTTCTTACCAAATTGTG....... 423
               :::   :::|||:::      ::::::   :::|||
    2309 uThrIleSerAspGluGlnArgSerAspAspPheThrValValAlaProG 2326

     424 .....AAAAGAAATAATTATAAAGCAGGGACTAACGGCCAT......... 459
           ::::::   |||   |||         |||||||||
    2326 lnProAsnHisAlaAsnThrLysIleTrpGlnAsnGlyHisIleAspIle 2342

     460 ...CCTTATGGCGGCGATTATCATATG..................... 483
           |||   :::   :::   |||:::
    2343 ThrProAsnAsnProSerGlyHisLeuIleAsnProThrGlnAlaMetAs 2359

     484 ....CCGCGTTTGCACAAATTTGTCACAGATGCAGAACCTGTTGAGATGA 529
                 :::|||:::   :::   ||||||      :::
    2359 pIleAlaTyrThrGluLysMetGlyAsnGlyAlaGluHisSerLysThrI 2376

     530 CCAGTTATATGGATGGG............TGGAAATACGCTGATTTAAAT 567
           :::   :::   |||            |||      |||      |||
    2376 leAsnValValArgGlyGlnAsnAsnGlnTrpThrIleAla......Asn 2390

     568 AAATACCCTGATCGTGTTCGAATCGGAGCA......GGCAGACAATATTG 611
         |||   ||||||   |||   :::   |||      |||:::      ::
    2391 Lys...ProAspTyrValThrLeuAspAlaHisThrGlyLysValThrPh 2406

     612 GCGGTCTGATGAAGACGAACCCAATAAC.................CGCG 643
         ::::::::::      :::||||||:::                  :::
    2406 eAsnAlaAsnThrValLysProAsnSerAlaIleThrIleThrProLysA 2423

     644 AAAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAAT 693
           ::::::   |||   ||||||:::                     |||
    2423 laGlyThrGlyHisSerAlaSerSer.....................Asn 2432

     694 ACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGA 743
               :::   ::::::   :::      |||||||||      :::||
    2433 ProSerThrLeuThrAlaProAlaThrHisThrValAsnThr...ThrGl 2448

     744 AAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTG 793
         |   :::|||         ||||||          |||:::::::::::::
    2448 uIleValLys......AspTyrGlySerAsnValThrAlaAlaGluIleA 2463

     794 GCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGG 843
         ::::::::                ::::      |||:::|||:::
    2463 snAsnAla..............ValGlnValAlaAsnLysArgThrAla 2474
```

```
 844 TTAATT...AATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAG 890
2475 ThrIleLysAsnGlyThrAlaMetProThrAsnLeuAlaGlyGlySerTh 2491

 891 CAATGGCTTC......CAGCTAGTTCGTAAAGATTGGTTCTATGATGAAA 934
2491 rThrThrIleProValThrValThrTyrAsnAspGlySerThrGluGluV 2508

 935 TCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAA...... 978
2508 al............GlnGluSerIlePheThrLysAlaAspLysArgGlu 2520

 979 ......AATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAA 1022
2521 LeuIleThrAlaLysAsnHisLeuAspAspProValSerThrAspGlyLy 2537

1023 A.......................ATCGATGCCAAACATAAACACTATT 1048
2537 sLysProGlyThrIleThrGlnTyrAsnAsnAlaIleHisAsnAlaGlnG 2554

1049 CTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCT 1098
2554 lnGlnIleAsnThrAlaLysThrGluAlaGlnGlnValIleAsnAsnGlu 2570

1099 TTATCC..............GAGACAGCAAGAGAACCTGTTTATCATGC 1133
2571 ArgAlaThrProGlnGlnValSerAspAlaLeuThrLysValArgAlaAl 2587

1134 TGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATA 1183
2587 aGlnThrLysIleAsnGluAlaLysAlaLeuLeuGlnAsnLysGluAspA 2604

1184 TTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATC 1233
2604 snSerGlnLeuValThrSerLysAsnAsn......LeuGlnSerServal 2618

1234 AACCAA......GGCGCGGGCGGTTTGTATTTTGAGGGT......AATTT 1271
2619 AsnGlnValProSerThrThrGlyMetThrGlnGlnSerIleAspAsnTy 2635

1272 TACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATA 1321
2635 rAsnAlaLysLysArgGluAlaGluThrGluIleThrAlaAlaGlnArgV 2652

1322 TCAGTGATGGCAGTACCGTTACTTGGAAA.................GTA 1353
2652 alIleAspAsnGlyAspAlaThrAlaGlnGlnIleSerAspGluAsnThr 2668

1354 AACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCT 1403
2669 ThrAlaGlnArgAsnAspLeuThrAsnGlnIleSerGlnAlaThrAsnLe 2685

1404 GGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTA 1453
2685 uAlaAlaValGluSerValLysGlnSerAlaAsnSerLeu...AspGlyA 2701

1454 AAGTC..................ATCTTAGATCAG.............. 1470
2701 laMetGlyAsnLeuGlnThrAlaIleAsnAspLysSerGlyThrLeuAla 2717

1471 ..............CAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAG 1505
2718 SerGlnAsnPheLeuAspAlaAspGluGln...LysArgAsnAlaTyrAs 2733

1506 TGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATA 1555
2733 nGln......AlaValSerAsnAlaGluThrIle...LeuAsnLysGlnT 2747

1556 ATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTG 1605
2747 hrGlyProAsnThrAlaLys..............ThrAlaValGluGln 2758

1606 GATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGA 1655
2759 AlaLeuAsnAsnValAsnSerAlaLysHisAlaLeuAsnGlyThrGln.. 2774

1656 AGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACC. 1704
2775 ................AsnLeuAsnAsnAlaLysGlnAlaAlaIleThrA 2786

1705 ..ATTACAGGCAATAAAGATATTACTACAACCGGCAATAACAACAACTTG 1752
2786 laIleAsnGlyAlaSerAspLeu.............AsnGlnLysGlnLys 2798

1753 GATAGCAAAAAAGAAATTGCC..................TACAACGGTTG 1784
2799 AspAlaLeuLysAlaGlnAlaAsnGlyAlaGlnArgValSerAsnAlaGl 2815

1785 GTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAATT 1834
2815 nAspValGlnArgAsnAlaThrGluLeuAsnThrAlaMetGly...ThrL 2831

1835 ACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAAT 1884
2831 euLysHisAlaIleAlaAspLysThrAsnThrLeuAlaSerSerLysTyr 2847

1885 TTAAACGGCAATATCACGCAA.......................ACAAA 1910
2848 ValAsnAlaAspSerThrLysGlnAsnAlaTyrThrThrLysValThrAs 2864

1911 CGGCAAACTGTTTTTCAGCGGCAGACCGACA..................... 1941
2864 nAlaGluHisIleIleSerGlyThrProThrValValThrThrProSerG 2881

1942 .....CCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAA 1986
2881 luValThrAlaAlaAlaAsnGlnValAsnSerAlaLysGlnGluLeuAsn 2897

1987 GGTATCCCACAAGGAGAAATCGTGTGGGAC......AACGATTGGATCGA 2030
2898 GlyAspGluArgLeuArgValAlaLysGlnAsnAlaAsnThrAlaIleAs 2914

2031 C..............CGCACATTTAAAGCGGAAAACTTCCATATTCAGG 2065
2914 pAlaLeuThrGlnLeuAsnThrProGlnLysAlaLysLeuLysGluGlnV 2931

2066 GCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGG 2115
2931 alGlyGlnAlaAsnArgLeuGluAspIleGlnThrValGlnThrAsnGly 2947

2116 CATTTAAGCAATCACGCC.............................. 2133
2948 GlnAlaLeuAsnAsnAlaMetLysGlyLeuArgAspSerIleAlaAsnGl 2964

2134 ..................CAAGCAGTTTTCGGTGTCGCACCGCATCAAA 2164
2964 uThrThrValLysAlaSerGlnAsnTyrThrAspAlaSerProAsnAsnG 2981

2165 GCCACACAATCTGTACACGT...TCGGACTGGACGGGTCTGACAAGTTGT 2211
2981 lnSerThrTyrAsnSerAlaValSerAsnAlaLysGlyIleIleAsnGln 2997

2212 ACCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTG........ 2253
2998 ThrAsnAsnProThrMetAspThrSerAlaIleThrGlnAlaThrThrGl 3014

2254 ..........AGCAAGACCGACATCAGAGGCAATGTCAGCCTT....... 2286
3014 nValAsnAsnAlaLysAsnGlyLeuAsnGlyAlaGluAsnLeuArgAsnA 3031

2287 ..GCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGC 2334
3031 laGlnAsnThrAlaLysGlnAsnLeuAsnThrLeuSerHisLeuThrAsn 3047

2335 AATCTTAGT.........................GCAGGCGGAGACAC 2357
```


3048 AsnGlnLysSerAlaIleSerSerGlnIleAspArgAlaGlyHisValSe 3064
2358 GCACTATACGGTTACGCGCAACGCC.....ACCCAAAACG 2392
3064 rGluValThrAlaAlaLysAsnAlaAlaThrGluLeuAsnThrGlnMetG 3081
2393 GCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACA 2442
3081 lyAsnLeuGluGlnAlaIleHisAspGlnAsnThrValLysGlnGlyVal 3097
2443 TTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAATCTAAGCAA 2492
3098 .....AsnPheThrAspAlaAspLysAlaLysArgAspAlaTyrTh 3111
2493 CAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAA 2542
3111 rAsnAlaValSerArgAlaGluThrIleLeuAsnLysThrGlnGlyAlaA 3128
2543 ACGTAAGC..... 2550
3128 snThrSerLysGlnAspValGluAlaAlaIleGlnAsnValThrSerAla 3144
2551 CATTCCGCACTCAACGGC.....AATGTCTCCCTAGCCGATAAGGCAGT 2594
3145 LysAsnAlaLeuAsnGlyAspGlnAsnValThrAsnAlaLysAsnThrAl 3161
2595 ATTCCATTTTGAAAACAGC.....C 2614
3161 aLysHisAlaLeuAsnAsnLeuThrSerIleAsnAsnAlaGlnLysArgA 3178
2615 GCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAA 2664
3178 spLeuThrThrLysIle.....AspGlnAlaThrThrValAla 3190
2665 GACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAAT.....TT 2708
3191 GlyValGluAlaValSerAsnThrGlyThrGlnLeuAsnThrAlaMetAl 3207
2709 AAACCTTGACAACGCC.....ACCATTACACTCAATTCCGCCT 2746
3207 aAsnLeuGlnAsnGlyIleAsnAspLysAlaAsnThrLeuAlaSerGluA 3224
2747 ATCGACACGATGCGGCAGGC.....GCGCAAACCGGCAGTGCG 2784
3224 snTyrHisAspAlaAspSerAspLysLysThrAlaTyrThrGlnAlaVal 3240
2785 GCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCC 2834
3241 ThrAsnAlaGluAsnIleLeuAsnLysAsnSer...GlySerAsnLeuAs 3256
2835 GCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACG..... 2871
3256 pLysAlaAlaValGluAsnAlaLeuSerGlnValThrAsnAlaLysGlyA 3273
2872 ..GTAAACGGC.....AAATTGAACGGTCAGGGA 2898
3273 laLeuAsnGlyAsnHisAsnLeuGluGlnAlaLysSerAsnAlaAsnThr 3289
2899 ACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAA 2948
3290 ThrIleAsnGlyLeuGlnHisLeuThrThrAlaGlnLysAspLysLeuLy 3306
2949 G.....CTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTG 2986
3306 sGlnGlnValGlnGlnAlaGlnAsnValAlaGlyValAspThrValLysS 3323
2987 TCAAC...AATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGTA 3033
3323 erSerAlaAsnThrLeuAsnGlyAlaMetGlyThrLeuArgAsnSerIle 3339
3034 GTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTC..... 3078
3340 .....GlnAspAsnThrAlaThrAsnAsnGlyGlnAsnTyrLeuAs 3353
3079 .....ACCCTGCAAAACGAACACGTCGATGCCGGCG 3109
3353 pAlaThrGluSerAsnLysThrAsnTyrAsnAsnAlaValAspSer.Ala 3369
3110 CATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAAT 3159
3370 AsnGlyValIleAsnAlaThrSerAsnProAsnMetAspAlaAsnAlaIl 3386
3160 CCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCGGGAGAAAC 3209
3386 e.AsnGlnIleAlaThrGlnValThrSerThrLysAsnAla..... 3399
3210 AGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCCAAACAACAGG 3259
3400 .....LeuAspGlyThrHisAsnLeuThr.GlnAlaLysGlnThrA 3413
3260 CGGAAAAA.....GACAACGCGCAAAGCCTT.....GAC 3288
3413 laThrAsnAlaIleAspGlyAlaThrAsnLeuAsnLysAlaGlnLysAsp 3429
3289 GCGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTGC 3338
3430 AlaLeuLysAlaGlnValThrSerAla...GlnArgValAlaAsnValTh 3445
3339 CGAACCGGCCCGGCAGGCAGGCGGGGAAAATGCCGGCATTATGCAGGCGG 3388
3445 rSerIleGlnGlnThrAlaAsnGluLeuAsnThrAlaMetGlyGlnLeuG 3462
3389 AGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAA 3438
3462 lnHisGlyIleAspAspGluAsnAlaThrLysGlnThr.....GlnLys 3476
3439 CAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCG 3488
3477 TyrArgAspAlaGlu...GlnSerLysLysThrAlaTyrAspGlnAla.V 3492
3489 CCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAAC 3538
3492 alAlaAlaAlaLysAlaIleLeuAsnLysGlnThrGlySerAsnSerAsp 3508
3539 CGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTT 3588
3509 LysAlaAlaValAspArgAlaLeuGlnGlnVal..... 3519
3589 TCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGT 3638
3520 .....ThrSerThrLysAspA 3525
3639 GTTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACA 3688
3525 laLeuAsnGlyAspAlaLysLeuAlaGluAlaLysAlaAlaAla..... 3539
3689 CCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGAC 3738
3540 .....LysGlnAsnLeuGlyThrLeuAsnHisIleTh 3550
3739 CTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCAT 3788
3550 rAsnAla.....GlnArgThrAlaLeuGluGly..... 3559
3789 CCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACGGCATCGGCA 3838
3560 .....GlnIleAsnGlnAlaThr 3565
3839 ACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGC 3888
3566 ThrValAspGlyValAsnThrValLysThrAsnAlaAsnThrLeuAspGl 3582
3889 AGGTTCGACATCGGCATCAGCG.....CGGGCGCGGGTTTTAG 3926
3582 yAlaMetAsnSerLeuGlnGlySerIleAsnAspLysAspAlaThrLeuA 3599

```
    3927 TAGCGGCAGCCTTTCAGACGGCATCAGAGGCAAAA...TCCGCCGCCGCG 3973
                  :::      |||:::|||:::|||   :::|||   :::
    3599 rgAsnGlnAsnTyrLeuAspAlaAspGluSerLysArgAsnAlaTyrThr 3615

    3974 TGCTGCATTACGGCATTCAGG.....................CAAGA 3999
         ::::::||||||   :::                          ::
    3616 GlnAlaValThrAlaAlaGluGlyIleLeuAsnLysGlnThrGlyGlyAs 3632

    4000 TACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGC...ACATCGGCGC 4046
         :|||:::::::::::::::|||:::|||   |||      |||   |||:
    3632 nThrSerLysAlaAspValAspAsnAlaLeuAsnThrValThrArgAlaL 3649

    4047 AACGCGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATA 4096
         ::   |||:::::::::   :::::::::   :::|||||||||||||
    3649 ysAlaAlaLeuAsnGlyAlaGluAsnLeuArgAsnThrLysThrSer... 3664

    4097 TCGCCACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCA 4146
                            :::|||   |||   ::::::|||   ::
    3665 .....................AlaThrAsnThrIleAsnGlyLeuProAs 3674

    4147 GATTATTCATTCAAACCGGCGCAAC 4171
         ::::   :::   ::::::   |||
    3674 nLeuThrGlnLeuGlnLysAspAsn 3682
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30293

seq_documentation_block:

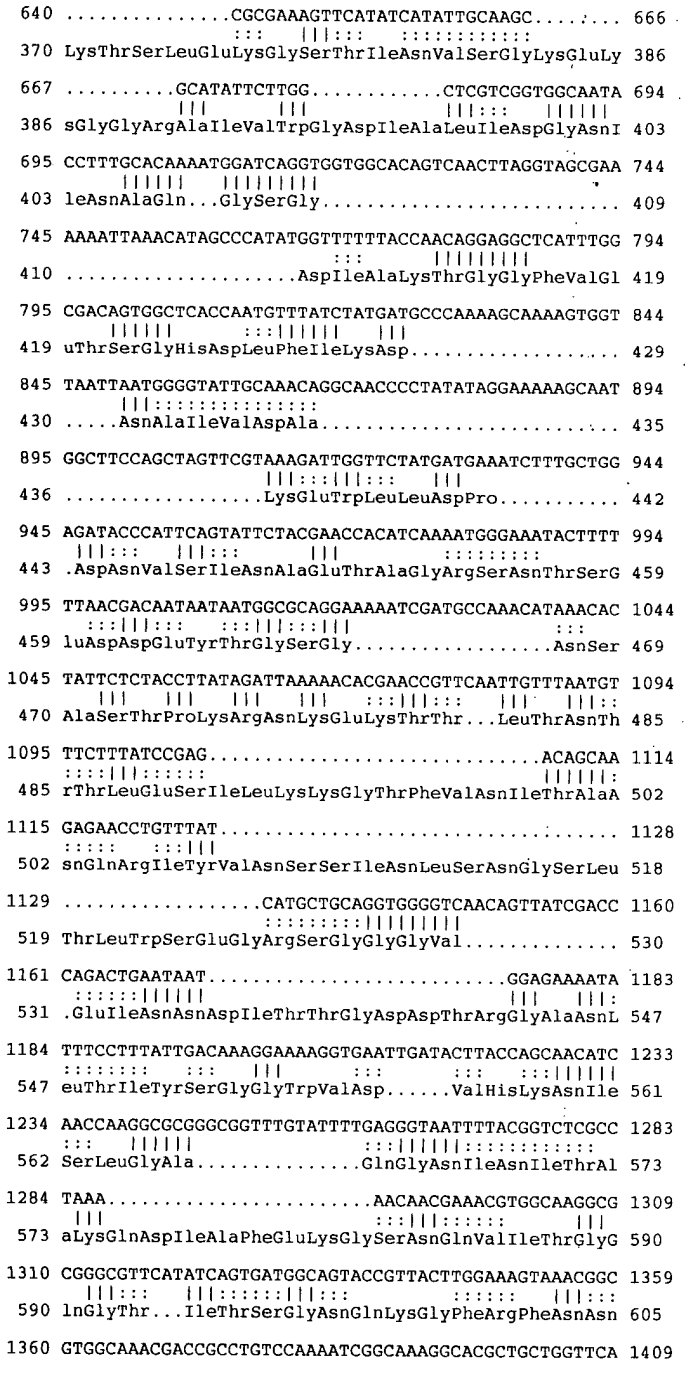
```
ID   AAW30293 standard; Protein; 1536 AA.
XX
AC   AAW30293;
XX
DT   14-APR-1998  (first entry)
XX
DE   Non-typeable Haemophilus high mol.wt. surface protein HMW1.
XX
KW   Non-typeable Haemophilus; high molecular weight surface protein;
KW   HMW1; hmw1A gene; immunogen; vaccine; otitis media.
XX
OS   Haemophilus influenzae strain 12.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 4
FT                   /note= "encoded by CTA"
FT   Misc-difference 98
FT                   /note= "encoded by GAT"
FT   Misc-difference 363
FT                   /note= "encoded by AAG"
XX
PN   WO9736914-A1.
XX
PD   09-OCT-1997.
XX
PF   01-APR-1997;   97WO-US04707.
XX
PR   01-APR-1996;   96US-0617697.
XX
PA   (BARE/) BARENKAMP S J.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1997-503038/46.
DR   N-PSDB; AAT90994 and AAT90996.
XX
PT   High molecular weight proteins of non-typeable Haemophilus
PT   influenzae - useful for vaccine production
XX
PS   Claim 7; Page 66-70; 183pp; English.
XX
CC   This protein comprises the high molecular weight surface protein
CC   HMW1 (125 kDa) of non-typeable Haemophilus influenzae strain 12 that
CC   has the immunological ability to protect against disease caused by a
CC   non-typeable Haemophilus strain and is characterised by at least
CC   one surface-exposed B-cell epitope that is recognised by monoclonal
CC   antibody AD6.  The HMW1 amino acid sequence was deduced from the
CC   hmw1 gene sequence (see AAT90994 and AAT90996).  The expressed protein
CC   is truncated, starting at residue 442 of the full-length gene
CC   product.  HMW2 (see AAW30294), HMW3 (see AAW30291) and HMW4 (see
CC   AAW30292) have also been identified.  A conjugate comprising HMW1
CC   linked to an antigen, hapten or polysaccharide, and a synthetic
CC   peptide of 6-150 amino acids corresponding to at least protective
CC   epitope of HMW1 are also claimed.  HMW proteins, conjugates and
CC   peptides can be used in vaccines, as immunogens for preparation of
CC   antibodies and as antigens for detection of these antibodies.
XX
SQ   Sequence   1536 AA;
```

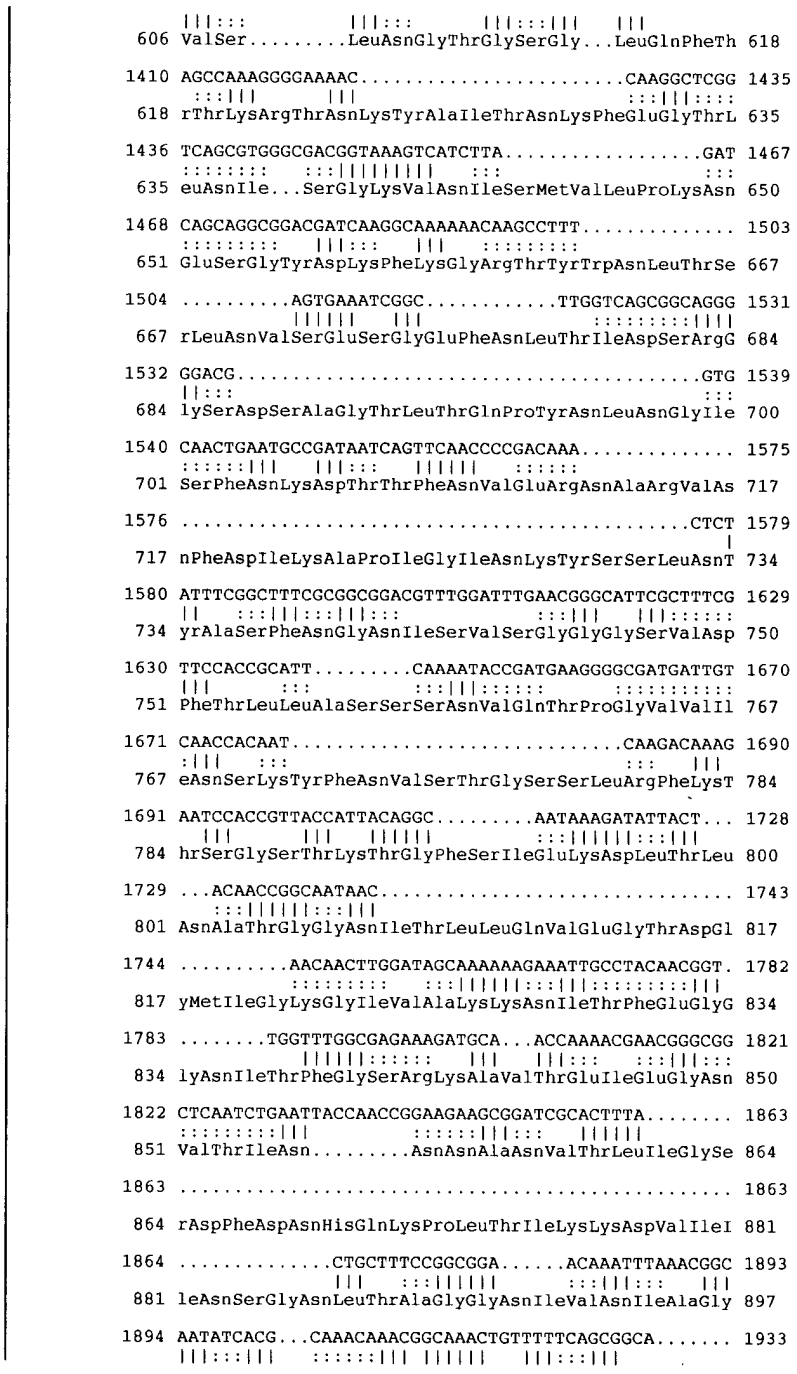
alignment_scores:

```
            Quality:  267.00              Length:   1297
              Ratio:  0.441                 Gaps:     73
 Percent Similarity:  46.646   Percent Identity: 21.126
```

alignment_block:

```
 US-09-303-518D-653 x AAW30293   ..

 Align seg 1/1  to: AAW30293  from: 1  to: 1536

    178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
        |||   |||   :::::|||||||   ||||||   :::|||
    199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal........ 211

    228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
              :::|||:::|||   ::::::   :::
    212 .......AsnLeuIleGlyGlyLysValLysAsnGlu.............. 221

    278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
          :::|||:::|||   ||||||      ::::::   |||||||||:::
    222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

    322 CAA.................................TATATTGTGAG 335
        :::                                 |||   :::::
    238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

    336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
        ::::   :::|||   :::   |||      |||:::   |||:::|||:
    254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271

    386 GCAAT.....................CCCGATCAGCACCGCTTTTCT 411
        ::|||                    :::|||   ::::::|||
    271 lyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287

    412 TACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCC 461
           :::   |||::::::::::      |||:::|||
    288 AlaAspSerValSerLysAsp......LysSerGly.............. 297

    462 TTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATG 511
                                              ::::::|||   :::|
    298 ..............................AsnIleValLeuSerA 303

    512 CAGAACCTGTTGAG.................ATGACCAGTTATATGGAT 543
        ||:::     |||                 :::::::::      :::
    303 laLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319

    544 GGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT......CG 587
              |||   :::   |||            |||:::|||      ::
    320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLy 336

    588 AATCGGAGCA.........................GGCAGACAATATT 610
        :   ||||||                        |||:::   |||
    336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353

    611 GGCGGTCTGATGAA...GACGAACCCAATAAC.................. 639
           :::||||||      |||   :::|||
    353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369
```





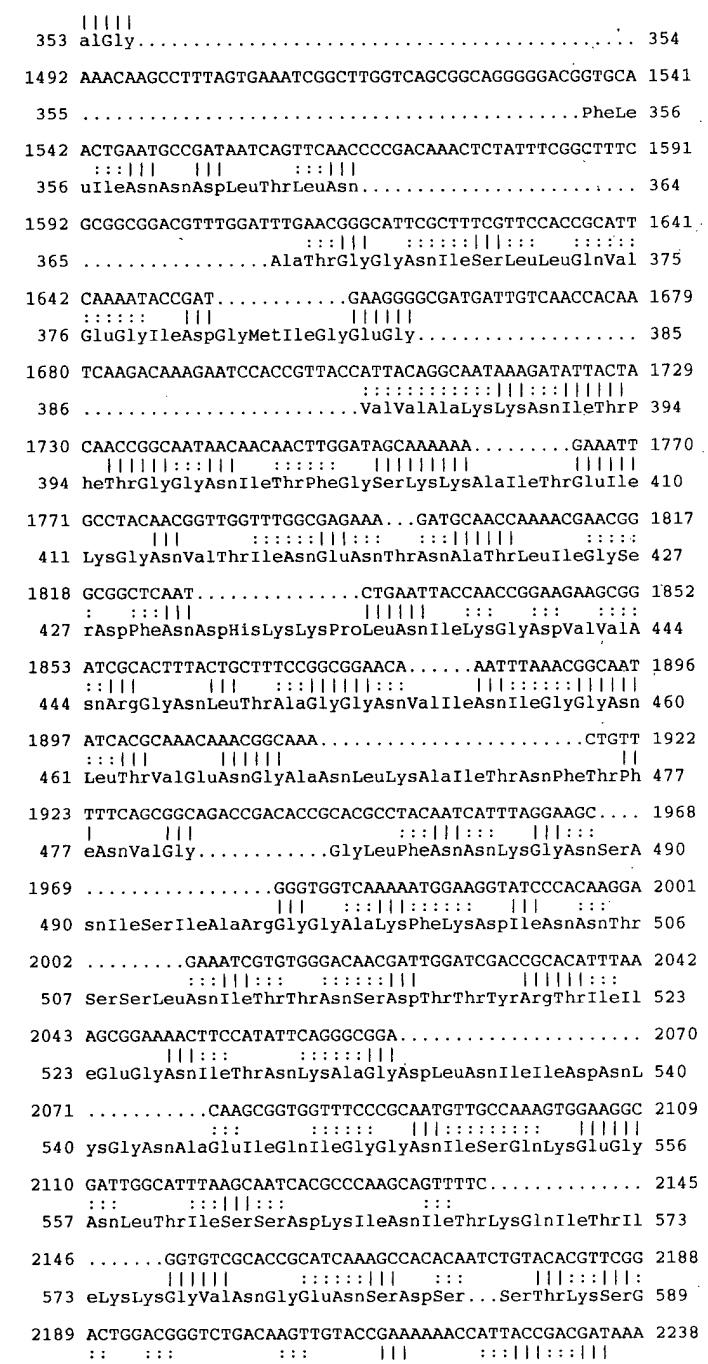
```
 898 AsnLeuThrValGluSerAsn.AlaAsn...PheLysAlaIleThrAsnP 913

1934 ........................GACCGACACCGCACGCCTACAATC 1957
                             |||:::      :::   :::|||
 913 heThrPheAsnValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIle 929

1958 ATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATC 2007
               ||||||   :::::|||        ::::::|||::
 930 SerIleAlaLysGlyGlyAlaArgPheLysAspIleAspAsnSerLysAs 946

2008 GTGTGGGACAACGATTGGATCGACCGC.ACATTTAAAGCGGAAAACTTCC 2056
     :   :::   |||   :::|||:::   |||::::::
 946 nLeuSerIleThrThrAsnSerSerSerThrTyrArg..... 958

2057 ATATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAA 2106
                    :::::::::|||   |||:::::::::   :::
 959 ................ThrIleIleSerGlyAsnIleThrAsnLysAsn 969

2107 GGCGATTGGCATTTAAGCAATCACGCC............CAAGCAGTTTT 2144
     ||||||   ::::::::::|||::::::           :::   :::
 970 GlyAspLeuAsnIleThrAsnGluGlySerAspThrGluMetGlnIleGl 986

2145 CGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGA 2194
      |||   :::   :::::::::   ::::::         ||||||
 986 yGlyAspValSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysI 1003

2195 CGGGTCTGACAAGTTGTACCGAAAAAACCATTACC...GACGATAAAGTG 2241
       ::::::|||            |||   ||||||            |||
1003 leAsnIleThr............LysGlnIleThrIleLysAlaGlyVal 1015

2242 ATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGA 2291
       ::::::   |||   :::|||      :::|||:::::::|||:::
1016 AspGlyGluAsnSerAspSerAspAlaThrAsnAsnAlaAsnLeuThrIl 1032

2292 TCACGCT.....................CATTTAAATCTCACAGGACTT. 2319
        :::                      ||||||:::::|||:::
1032 eLysThrLysGluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPheA 1049

2320 ..................................GCCACACTCAACGGC 2334
                                        |||::::::::::
1049 snLysAlaGluIleThrAlaLysAspGlySerAspLeuThrIleGlyAsn 1065

2335 AATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCAC 2384
     :::   ||||||   |||   :::       ||||||   |||   ::
1066 ThrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGlnVa 1082

2385 CCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATC 2434
     ::::::::::::::::|||      |||:::      :::      ::::
1082 lLysAspSerLysIleSerAlaAspGlyHisLysValThrLeuHisSerL 1099

2435 AAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAAT 2484
     :::::      :::|||:::::::::   :::::::::::|||   |||
1099 ysValGluThrSerGlySerAsnAsnAsnThrGluHisSerSerAspAsn 1115

2485 CTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGC 2534
        ::::::   :::::::::      ::::::|||:::::::::|||
1116 AsnAlaGlyLeuThrIleAspAlaLysAsnValThrValAsnAsnAsnIl 1132

2535 TAAGGCAAAC......GTAAGCCATTCCGCACTCAACGGCAATGTC.... 2574
        ::::::      ||||||   ||||||   :::|||::::::
1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyGluIleThrT 1149

2575 ...........TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGC 2613
                :::      :::   ::::::   :::   |||   :::
1149 hrLysThrGlyThrThrIleAsnAlaThrThrGlyAsnValGluIleThr 1165

2614 CGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACG.........GCATT 2654
           ||||||:::|||   ||||||   ::::::         :::::
1166 AlaGlnThrGlySerIleLeuGlyGlyIleGluSerSerSerGlySerVa 1182

2655 ACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCA 2704
     :   |||      :::|||   :::|||                ::::::|
1182 lThrLeuThrAlaThrGluGlyAlaLeu............AlaValSerA 1195

     2705 ATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACAC 2754
          ||::::::   :::::::::|||:::|||   ||||||
     1195 snIleSerGlyAsnThrValThrValThrAlaAsnSer..... 1207

     2755 GATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCG 2804
                  ||||||   |||      |||::::::
     1208 .........GlyAlaLeuThrThrLeuAlaGlySer..... 1216

     2805 TTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCC 2854
           :::   ::::::      |||||||||      ||||||:::::|||
     1217 .ThrIleLysGlyThrGluSerValThr......ThrSerSerGlnSerG 1231

     2855 GTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA... 2901
                          ::::::|||   ::::::|||   ||||||
     1231 ly............AspIleGlyGlyThrIleSerGly...GlyThrVal 1242

     2902 ...TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAA 2948
               :::      :::::::|||            ::::::|||:::||
     1243 GluValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLy 1259

     2949 GCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTC.......... 2988
          |   |||   ::::::|||::::::   |||   |||:::
     1259 s...AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleG 1275

     2989 ..............AACAATACCGGCAACGAACCCGTAAGTCTCGAGCAA 3024
                         :::||||||   |||      ::::::      :::
     1275 lyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1291

     3025 TTGACGGTA.......................GTGGAAGGAAAAGACAA 3050
          |||||||||                      :::||||||      ::
     1292 LeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaTh 1308

     3051 CACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCG 3100
          :         |||::::::      |||      ::::::|||:::
     1308 rLeuThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleT 1325

     3101 ATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGC 3150
            ::: |||   |||   ::::::|||||||||:::   :::|||   ||
     1325 hrSer.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAl 1341

     3151 CTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACA 3187
          |      |||::::::   |||   ::::::   |||
     1341 aGlySerIleAsnAlaAlaAsnValThrLeuAsnThr 1353
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01845

seq_documentation_block:

```
ID   AAB01845 standard; Protein; 1180 AA.
XX
AC   AAB01845;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain 15 mature HMW2A protein, SEQ ID NO:65.
XX
KW   Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain 15.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
```

```
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52194.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 8; Fig 27A-R; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   1180 AA;


alignment_scores:
            Quality:  266.00              Length:   1110
              Ratio:  0.494                Gaps:     57
 Percent Similarity:  48.468   Percent Identity: 21.261

alignment_block:
 US-09-303-518D-653 x AAB01845   ..

 Align seg 1/1  to: AAB01845  from: 1  to: 1180

     574 CCTGATCGTGTTCGAATCGGAGCAGGC..................... 600
         |||:::::::|||   |||:::||||||
       1 ProGluAsnValTyrIleAsnAlaGlyAspAlaGlyArgSerAspThrAs 17

     601 .........AGACAATATTGGCGGTCTGATGAA......GACGAACCC. 633
                 :::::::|||      :::   |||      |||   |||
      17 nLeuGluAsnGluGluTyrThrGlyThrGlyGluSerAlaAspThrProL 34

     634 .....AATAACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGG 678
             |||||||::   ::::::           |||:::   :::
      34 ysArgAsnAsnAsnThrLysThrThrLeuThrAsnSerThrLeuGluLys 50

     679 CTCGTCGGTGGCAATACCTTTGCACAA..................... 705
         :::::::::   ::::::|||::::::
      51 IleLeuAlaArgGlySerPheValAsnIleThrAlaAsnAsnGluIleAr 67

     706 ..............AATGGATCAGGTGGTGGCACAGTCAACTTAGGTA 739
                       |||   :::|||::::::   ::::::|||   |
      67 gValAsnSerAspIleAsnIleGlyGlyAsnSerHisLeuThrLeuTrpS 84

     740 GCGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCA 789
         ||:::|||   |||:::|||      |||   |||   :::|||:::
      84 erSerLysAsnLysAsnSer......GlyValLeuIleAsnGlyAsnIle 98

     790 TTTGGCGACAGTGGCTCACCAATGTTTATCTAT..............GA 824
            :::   :::::::::   :::   ||||||               ||
      99 ThrSerThrAlaAsnGlyAsnLeuThrIleTyrSerSerGlyTrpValAs 115

     825 TGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTG.........C 862
         |   :::|||:::      |||   :::|||   |||         :
     115 pIleHisLysAsnIleThrLeuGluSerGlyArgLeuAsnIleThrThrL 132

     863 AAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTAGTTCGT 912
         ::   |||:::      :::   |||:::|||:::::   :::::
     132 ysGluGlyAspValAlaPheGluLysGlyAsnAsnLeuThrIleThrGly 148

     913 AAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATT 962
         :::                 |||   ||||||::::::   :::
     149 Gln...........GlyThrIleThrAlaGlyAsnAsnLysGlyPheAr 161

     963 CTACGAA..................................... 969
          :::|||
     161 gPheGluAsnValSerLeuAsnGlyThrGlyThrGlyLeuLeuPheAsnL 178

     970 ........CCACATCAAAATGGGAAATAC.........TTTTTTAACGAC 1002
                 |||:::::::|||:::::            :::||||||
     178 euSerArgProGlnLysAsnAsnSerLeuValThrAsnTyrPheAsnGly 194

    1003 AATAATAATGGCGCAGGAAAAATC.....................GATGC 1031
         :::   |||   :::|||:::::                      :::||
     195 ThrLeuAsnIleSerGlySerValAsnIleSerMetIleProProAsnAl 211

    1032 CAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTT. 1080
         |      :::   ||||||          |||   |||   ||||||
     211 aThrSerAsnTrpTyrSer.........ArgTyrLysGlyArgThrTyrT 225

    1081 ..............................CAATTGTTTAATGTTTCT 1098
                                       :::   ||||||::::::
     225 rpAsnIleThrHisLeuAsnAlaSerGluAspSerAsnPheAsnLeuThr 241

    1099 TTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAA 1148
         ::::::::::::|||::::::          :::|||      :::::
     242 IleAspSerSerAlaGluAsp.........GlySerAlaProLeuLeuSe 255

    1149 CAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACA 1198
         :||||||          ||||||          :::||||||||||
     255 rSerTyr......ThrLeuAsn.........GlyIleSerPheThrThrA 267

    1199 AAGGAAAAGGTGAATTGATACTTACCAGCAACATCAACCAAGGCGCGGGC 1248
                                 |||   |||:::|||::::::|||
     267 sp..................ThrThrPheAsnValAsnLysAsnAla... 276

    1249 GGTTTGTATTTTGAGGGTAATTTT.....................ACGGT 1277
                    :::   ||||||                      |||::
     277 ............LysValAsnPheAsnIleLysAlaProIleGlyThrIl 289

    1278 CTCGCCTAAAAACAACGAAACGTGGCAA......GGCGCGGGCGTTCATA 1321
         ::::      ||||||   ::::::          ::::::::::::   :
     289 eAsnGlnTyrAsnAsnLeuAsnTyrAlaLeuPheAsnGlyAsnIleServ 306

    1322 TCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGC............. 1359
         ::|||   |||::::::||||||:::::::::|||:::
     306 alSerGlyGlyGlyAsnValThrPheArgLeuAsnAlaSerSerSerAsn 322

    1360 ..................GTGGCAAACGACCGCCTGTCCAAAATCGGCAA 1391
                           :::   |||::::::      :::   :::||
     323 GlnGlnThrProGlyValIleIleAsnSerLysHisLeuAsnAlaSerLy 339

    1392 AGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCG 1441
         ||||:::   |||         :::   |||:::   ||||||:::::|
     339 sGlySerSerLeu.........ArgPheGluThrThrGlySerThrLysV 353

    1442 TGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAA 1491
```


353 alGly 354
1492 AAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCA 1541
355 PheLe 356
1542 ACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTC 1591
356 uIleAsnAsnAspLeuThrLeuAsn 364
1592 GCGGCCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATT 1641
365 AlaThrGlyGlyAsnIleSerLeuLeuGlnVal 375
1642 CAAAATACCGAT GAAGGGGCGATGATTGTCAACCACAA 1679
376 GluGlyIleAspGlyMetIleGlyGluGly 385
1680 TCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACTA 1729
386 ValValAlaLysLysAsnIleThrP 394
1730 CAACCGGCAATAACAACAACTTGGATAGCAAAAAA GAAATT 1770
394 heThrGlyGlyAsnIleThrPheGlySerLysLysAlaIleThrGluIle 410
1771 GCCTACAACGGTTGGTTTGGCGAGAAA GATGCAACCAAAACGAACGG 1817
411 LysGlyAsnValThrIleAsnGluAsnThrAsnAlaThrLeuIleGlySe 427
1818 GCGGCTCAAT CTGAATTACCAACCGGAAGAAGCGG 1852
427 rAspPheAsnAspHisLysLysProLeuAsnIleLysGlyAspValValA 444
1853 ATCGCACTTTACTGCTTTCCGGCGGAACA AATTTAAACGGCAAT 1896
444 snArgGlyAsnLeuThrAlaGlyGlyAsnValIleAsnIleGlyGlyAsn 460
1897 ATCACGCAAACAAACGGCAAA CTGTT 1922
461 LeuThrValGluAsnGlyAlaAsnLeuLysAlaIleThrAsnPheThrPh 477
1923 TTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGC 1968
477 eAsnValGly GlyLeuPheAsnAsnLysGlyAsnSerA 490
1969 GGGTGGTCAAAAATGGAAGGTATCCCACAAGGA 2001
490 snIleSerIleAlaArgGlyGlyAlaLysPheLysAspIleAsnAsnThr 506
2002 GAAATCGTGTGGGACAACGATTGGATCGACCGCACATTTAA 2042
507 SerSerLeuAsnIleThrThrAsnSerAspThrThrTyrArgThrIleIl 523
2043 AGCGGAAAACTTCCATATTCAGGGCGGA 2070
523 eGluGlyAsnIleThrAsnLysAlaGlyAspLeuAsnIleIleAspAsnL 540
2071 CAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGC 2109
540 ysGlyAsnAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysGluGly 556
2110 GATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTC 2145
557 AsnLeuThrIleSerSerAspLysIleAsnIleThrLysGlnIleThrIl 573
2146 GGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGG 2188
573 eLysLysGlyValAsnGlyGluAsnSerAspSer SerThrLysSerG 589
2189 ACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGACGATAAA 2238
589 lnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrGlnAspLeu 605
2239 GTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGC 2288
606 AsnIleSerGlyPheAsnLysAlaLysIleValAlaLysAspSer 620
2289 CGATCACGCTCATTTAAATCTCACA GGACTTGCCACACTCAACGGCA 2335
621 SerAsnLeuThrIleGlyAsnSerAspAspSerGlyA 633
2336 ATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCACC 2385
633 snThrSerAla LysThrValThrPheAsnAsnVal 644
2386 CAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCA 2435
645 LysAspSerLysIleSerAlaAspGlyHis LysValThrLeuAsnSe 660
2436 A GCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCAT 2479
660 rLysValLysThrLeuSerAspAsnAspAsnAsnThrGluGlyGlySerA 677
2480 TTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGAC 2529
677 spAsnAsnThrGlyLeuThrIleThrAlaLysAspValGluValAsnAsn 693
2530 AACGCTAAGGCAAAC GTAAGCCATTCCGCACTCAACGGCAATGT 2573
694 AsnIleThrSerHisLysThrValAsnValSerAlaAlaAsnGlyGlyIl 710
2574 C TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAA 2608
710 eThrThrLysThrGlyThrThrIleAsnAlaThrAlaGlyAsnValGluI 727
2609 ACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACAC 2658
727 leThrAlaHisThrGlySerIleGlnGlyGly IleGlu 739
2659 TTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAA TT 2699
740 SerLysProGlySerValThrIleValAlaGlyGlyAspThrLeuAlaVa 756
2700 AGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATC 2749
756 lGlyAsnIleSerGlyAsnAlaValThrValThrAlaAsnSer 770
2750 GACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGC 2799
771 GlyAlaLeuThrThrLeuAlaGlySer 779
2800 CGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGA 2849
780 ThrIleLysGlyThrGluSerIleThr ThrSerSerGl 792
2850 ATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGA 2898
792 nSerGly AsnIleGlyGlyLysIleSerGlyLysThrV 805
2899 ACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTG 2946
805 alAsnValLysAlaThrAsnSerLeuThrThrGlnAlaAspSerLysIle 821
2947 AAGCTGGCGGAAAGT 2961
822 GluAlaThrGluGlyGluAlaAsnValThrSerLysThrSerIleIleGl 838
2962 TCCGAAGGCACTTACACCTTGGCTGTCAACAATACCGGCA 3001
838 yGlyThrIleSerGlyGlyThr ValGluValThrAlaT 851
3002 ACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAAC 3051
851 hrGluGlyLeuThrThrGlnAlaGlySerThrIleThrGlyThrGluSer 867

```
3052 ACACCGCTGTCCGAA.........AATCTTAATTTCACCCTGCAAAACGA 3092
     :::      |||:::         |||::::::      :::::::::
 868 ValThrThrSerSerGlnSerGlyAsnIleGlyGlyMetIleSerGlyGl 884

3093 ACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCG 3142
         |||::::::::::|||   :::   :::|||||   |||:::|||
 884 yLysValGluValSerAlaThrLys...AspLeuIleThrLysSerGly. 899

3143 AGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTC 3192
                                 :::|||::::::
 900 ............................SerGluIleLysAlaThrAla 906

3193 GGCAAGGCGGGAGAAACAGAGGCCGCC..................... 3219
     |||::::::::::   |||:::|||:::
 907 GlyGluValAsnValThrSerAlaThrGlyThrIleAspGlyThrIleSe 923

3220 ..............TTGACGGCAAAACAGGCACAACTTGCCGCCAAAC 3253
                    :::|||||||:::   :::::::|||:::::::::
 923 rGlyAsnThrValAsnValThrAlaAsnThrGlyAspLeuThrValGluA 940

3254 AACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCC 3303
     ::   |||:::   |||   :::   :::      :::|||   |||:::
 940 spAlaAlaLysIleAspAlaThrGlyGlyAlaAlaThrLeuThrAlaThr 956

3304 GGGCGCAATGCCACCGAAAAGGCAGAAAGT 3333
     :::   :::   |||   ||||||:::|||
 957 SerGlyLysLeuThrThrLysAlaSerSer 966
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01844

seq_documentation_block:

ID AAB01844 standard; Protein; 1188 AA.
XX
AC AAB01844;
XX
DT 11-SEP-2000 (first entry)
XX
DE Haemophilus influenzae strain 15 HMW2A protein, SEQ ID NO:63.
XX
KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
OS Haemophilus influenzae strain 15.
XX
PN WO200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
DR WPI; 2000-303789/26.
DR N-PSDB; AAA52193.
XX
PT Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
PS Claim 12; Fig 27A-R; 307pp; English.
XX
CC The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMWA
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typeable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents an HMWA protein from a non-typeable strain of
CC H. influenzae.
XX
SQ Sequence 1188 AA;

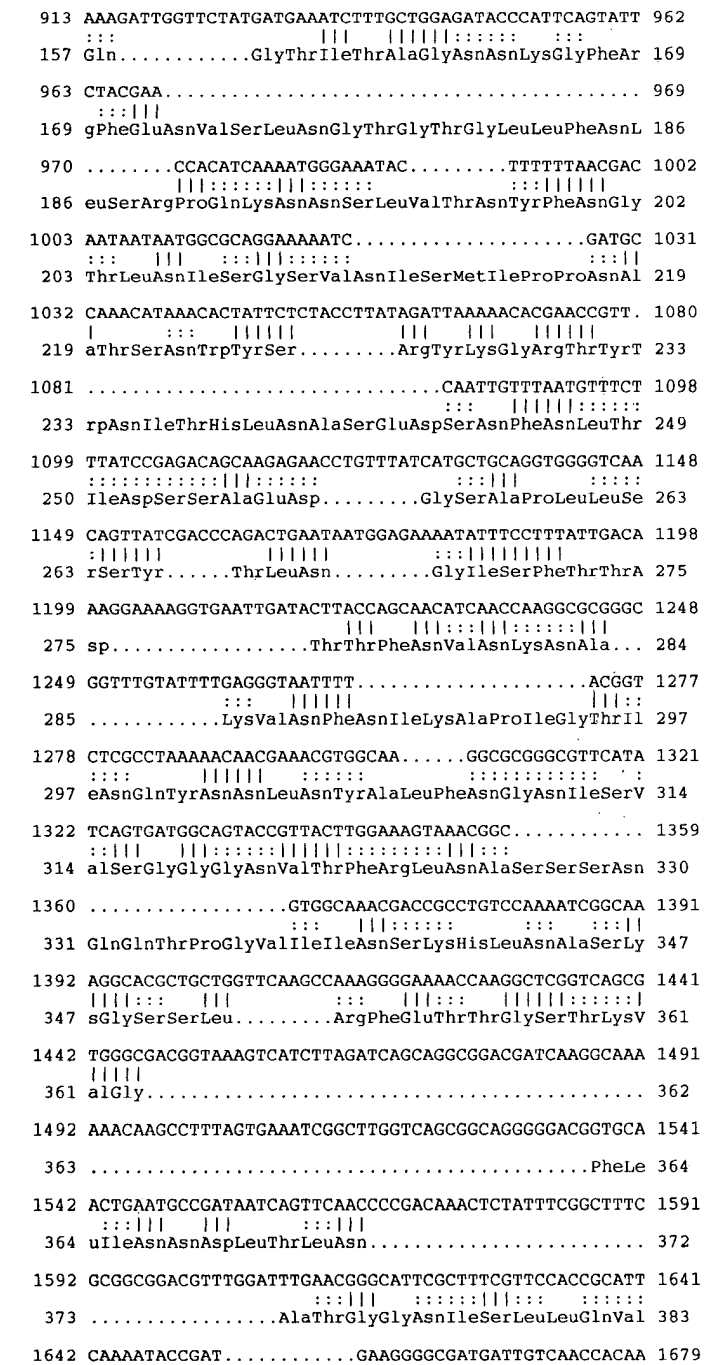
alignment_scores:

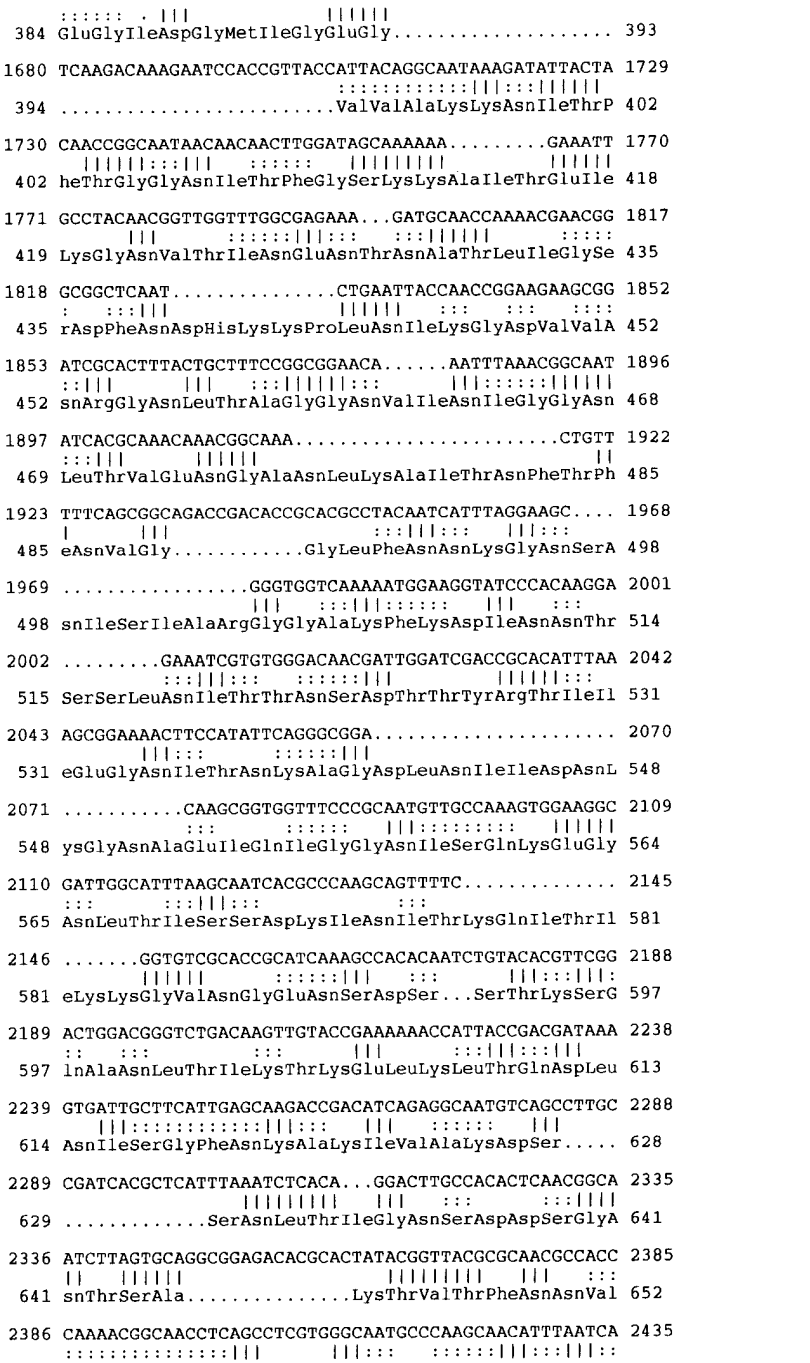
Quality:	266.00	Length:	1110
Ratio:	0.494	Gaps:	57
Percent Similarity:	48.468	Percent Identity:	21.261

alignment_block:
US-09-303-518D-653 x AAB01844 ..

Align seg 1/1 to: AAB01844 from: 1 to: 1188

```
 574 CCTGATCGTGTTCGAATCGGAGCAGGC..................... 600
     |||::::::|||   |||:::||||||
   9 ProGluAsnValTyrIleAsnAlaGlyAspAlaGlyArgSerAspThrAs 25

 601 .........AGACAATATTGGCGGTCTGATGAA......GACGAACCC. 633
             ::::::|||      :::   |||      |||   |||
  25 nLeuGluAsnGluGluTyrThrGlyThrGlyGluSerAlaAspThrProL 42

 634 .....AATAACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGG 678
          ||||||:::   ::::::           |||:::   :::
  42 ysArgAsnAsnAsnThrLysThrThrLeuThrAsnSerThrLeuGluLys 58

 679 CTCGTCGGTGGCAATACCTTTGCACAA..................... 705
     :::::::::   ::::::|||::::::
   59 IleLeuAlaArgGlySerPheValAsnIleThrAlaAsnAsnGluIleAr 75

 706 ..............AATGGATCAGGTGGTGGCACAGTCAACTTAGGTA 739
                   |||   :::|||::::::   ::::::|||   |
   75 gValAsnSerAspIleAsnIleGlyGlyAsnSerHisLeuThrLeuTrpS 92

 740 GCGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCA 789
     ||:::|||   |||:::|||      |||   |||   :::|||:::
  92 erSerLysAsnLysAsnSer......GlyValLeuIleAsnGlyAsnIle 106

 790 TTTGGCGACAGTGGCTCACCAATGTTTATCTAT..............GA 824
        :::   :::::::::   :::   ||||||               ||
 107 ThrSerThrAlaAsnGlyAsnLeuThrIleTyrSerSerGlyTrpValAs 123

 825 TGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTG...........C 862
     |   :::|||:::      |||   :::|||   |||           :
 123 pIleHisLysAsnIleThrLeuGluSerGlyArgLeuAsnIleThrThrL 140

 863 AAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTAGTTCGT 912
     ::   |||:::     :::   |||:::|||::::::   ::::::
 140 ysGluGlyAspValAlaPheGluLysGlyAsnAsnLeuThrIleThrGly 156
```


913 AAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATT 962
157 Gln.....GlyThrIleThrAlaGlyAsnAsnLysGlyPheAr 169
963 CTACGAA..... 969
169 gPheGluAsnValSerLeuAsnGlyThrGlyThrGlyLeuLeuPheAsnL 186
970 .....CCACATCAAAATGGGAAATAC.....TTTTTTAACGAC 1002
186 euSerArgProGlnLysAsnAsnSerLeuValThrAsnTyrPheAsnGly 202
1003 AATAATAATGGCGCAGGAAAAATC.....GATGC 1031
203 ThrLeuAsnIleSerGlySerValAsnIleSerMetIleProProAsnAl 219
1032 CAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTT. 1080
219 aThrSerAsnTrpTyrSer.....ArgTyrLysGlyArgThrTyrT 233
1081 .....CAATTGTTTAATGTTTCT 1098
233 rpAsnIleThrHisLeuAsnAlaSerGluAspSerAsnPheAsnLeuThr 249
1099 TTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAA 1148
250 IleAspSerSerAlaGluAsp.....GlySerAlaProLeuLeuSe 263
1149 CAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACA 1198
263 rSerTyr.....ThrLeuAsn.....GlyIleSerPheThrThrA 275
1199 AAGGAAAAGGTGAATTGATACTTACCAGCAACATCAACCAAGGCGCGGGC 1248
275 sp.....ThrThrPheAsnValAsnLysAsnAla... 284
1249 GGTTTGTATTTTGAGGGTAATTTT.....ACGGT 1277
285 .....LysValAsnPheAsnIleLysAlaProIleGlyThrIl 297
1278 CTCGCCTAAAAACAACGAAACGTGGCAA.....GGCGCGGGCGTTCATA 1321
297 eAsnGlnTyrAsnAsnLeuAsnTyrAlaLeuPheAsnGlyAsnIleSerV 314
1322 TCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGC..... 1359
314 alSerGlyGlyGlyAsnValThrPheArgLeuAsnAlaSerSerSerAsn 330
1360 .....GTGGCAAACGACCGCCTGTCCAAAATCGGCAA 1391
331 GlnGlnThrProGlyValIleIleAsnSerLysHisLeuAsnAlaSerLy 347
1392 AGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCG 1441
347 sGlySerSerLeu.....ArgPheGluThrThrGlySerThrLysV 361
1442 TGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAA 1491
361 alGly..... 362
1492 AAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCA 1541
363 .....PheLe 364
1542 ACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTC 1591
364 uIleAsnAsnAspLeuThrLeuAsn..... 372
1592 GCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATT 1641
373 .....AlaThrGlyGlyAsnIleSerLeuLeuGlnVal 383
1642 CAAAATACCGAT.....GAAGGGGCGATGATTGTCAACCACAA 1679


384 GluGlyIleAspGlyMetIleGlyGluGly..... 393
1680 TCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACTA 1729
394 .....ValValAlaLysLysAsnIleThrP 402
1730 CAACCGGCAATAACAACAACTTGGATAGCAAAAAA.....GAAATT 1770
402 heThrGlyGlyAsnIleThrPheGlySerLysLysAlaIleThrGluIle 418
1771 GCCTACAACGGTTGGTTTGGCGAGAAA...GATGCAACCAAAACGAACGG 1817
419 LysGlyAsnValThrIleAsnGluAsnThrAsnAlaThrLeuIleGlySe 435
1818 GCGGCTCAAT.....CTGAATTACCAACCGGAAGAAGCGG 1852
435 rAspPheAsnAspHisLysLysProLeuAsnIleLysGlyAspValValA 452
1853 ATCGCACTTTACTGCTTTCCGGCGGAACA.....AATTTAAACGGCAAT 1896
452 snArgGlyAsnLeuThrAlaGlyGlyAsnValIleAsnIleGlyGlyAsn 468
1897 ATCACGCAAACAAACGGCAAA.....CTGTT 1922
469 LeuThrValGluAsnGlyAlaAsnLeuLysAlaIleThrAsnPheThrPh 485
1923 TTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGC.... 1968
485 eAsnValGly.....GlyLeuPheAsnAsnLysGlyAsnSerA 498
1969 .....GGGTGGTCAAAAATGGAAGGTATCCCACAAGGA 2001
498 snIleSerIleAlaArgGlyGlyAlaLysPheLysAspIleAsnAsnThr 514
2002 .....GAAATCGTGTGGGACAACGATTGGATCGACCGCACATTTAA 2042
515 SerSerLeuAsnIleThrThrAsnSerAspThrThrTyrArgThrIleIl 531
2043 AGCGGAAAACTTCCATATTCAGGGCGGA..... 2070
531 eGluGlyAsnIleThrAsnLysAlaGlyAspLeuAsnIleIleAspAsnL 548
2071 .....CAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGC 2109
548 ysGlyAsnAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysGluGly 564
2110 GATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTC..... 2145
565 AsnLeuThrIleSerSerAspLysIleAsnIleThrLysGlnIleThrIl 581
2146 .....GGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGG 2188
581 eLysLysGlyValAsnGlyGluAsnSerAspSer...SerThrLysSerG 597
2189 ACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGACGATAAA 2238
597 lnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrGlnAspLeu 613
2239 GTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGC 2288
614 AsnIleSerGlyPheAsnLysAlaLysIleValAlaLysAspSer..... 628
2289 CGATCACGCTCATTTAAATCTCACA...GGACTTGCCACACTCAACGGCA 2335
629 .....SerAsnLeuThrIleGlyAsnSerAspAspSerGlyA 641
2336 ATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCACC 2385
641 snThrSerAla.....LysThrValThrPheAsnAsnVal 652
2386 CAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCA 2435

```
 653 LysAspSerLysIleSerAlaAspGlyHis...LysValThrLeuAsnSe 668

2436 A......GCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCAT 2479
     :        ||||||:::   |||   :::   :::::::::::|||
 668 rLysValLysThrLeuSerAspAsnAspAsnAsnThrGluGlyGlySerA 685

2480 TTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGAC 2529
       |||   ::::::   ::::::         ::::::   :::::::::
 685 spAsnAsnThrGlyLeuThrIleThrAlaLysAspValGluValAsnAsn 701

2530 AACGCTAAGGCAAAC......GTAAGCCATTCCGCACTCAACGGCAATGT 2573
     |||      ::::::      |||:::   ||||||   ||||||:::::
 702 AsnIleThrSerHisLysThrValAsnValSerAlaAlaAsnGlyGlyIl 718

2574 C..............TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAA 2608
     :              :::      :::   ::::::   :::   |||
 718 eThrThrLysThrGlyThrThrIleAsnAlaThrAlaGlyAsnValGluI 735

2609 ACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACAC 2658
       :::      ||||||:::|||:::||||||             ::::::
 735 leThrAlaHisThrGlySerIleGlnGlyGly............IleGlu 747

2659 TTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAA.........TT 2699
        |||   ::::::   |||:::   :::|||   :::        ::
 748 SerLysProGlySerValThrIleValAlaGlyGlyAspThrLeuAlaVa 764

2700 AGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATC 2749
     :||||||::::::   :::   :::|||:::|||   ||||||
 764 lGlyAsnIleSerGlyAsnAlaValThrValThrAlaAsnSer..... 778

2750 GACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGC 2799
                |||||   |||      |||::::::
 779 .............GlyAlaLeuThrThrLeuAlaGlySer......... 787

2800 CGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGA 2849
           :::   ::::::      |||:::|||      ||||||:::::
 788 ......ThrIleLysGlyThrGluSerIleThr......ThrSerSerGl 800

2850 ATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGA. 2898
     :|||              :::::::::||||||:::::::|||:::
 800 nSerGly............AsnIleGlyGlyLysIleSerGlyLysThrV 813

2899 ..ACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTG 2946
       :::   :::      ::::::|||            ::::::|||:::
 813 alAsnValLysAlaThrAsnSerLeuThrThrGlnAlaAspSerLysIle 829

2947 AAGCTGGCGGAAAGT................................. 2961
     :::   :::|||:::
 830 GluAlaThrGluGlyGluAlaAsnValThrSerLysThrSerIleIleGl 846

2962 ..........TCCGAAGGCACTTACACCTTGGCTGTCAACAATACCGGCA 3001
               |||   ||||||            |||:::   |||::::
 846 yGlyThrIleSerGlyGlyThr............ValGluValThrAlaT 859

3002 ACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAAC 3051
     ::|||   ::::::   :::      :::::::::   |||   ::::::
 859 hrGluGlyLeuThrThrGlnAlaGlySerThrIleThrGlyThrGluSer 875

3052 ACACCGCTGTCCGAA.........AATCTTAATTTCACCCTGCAAAACGA 3092
     :::      |||:::         |||::::::      :::::::::
 876 ValThrThrSerSerGlnSerGlyAsnIleGlyGlyMetIleSerGlyGl 892

3093 ACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCG 3142
        |||:::::::::|||   :::   :::||||||   |||:::|||
 892 yLysValGluValSerAlaThrLys...AspLeuIleThrLysSerGly. 907

3143 AGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTC 3192
                                 :::|||::::::
 908 ............................SerGluIleLysAlaThrAla 914

3193 GGCAAGGCGGGAGAAACAGAGGCCGCC..................... 3219
     |||:::::::::   |||:::|||:::
 915 GlyGluValAsnValThrSerAlaThrGlyThrIleAspGlyThrIleSe 931

     3220 ...............TTGACGGCAAAACAGGCACAACTTGCCGCCAAAC 3253
                           :::|||||||:::   :::::|||:::::::::::
      931 rGlyAsnThrValAsnValThrAlaAsnThrGlyAspLeuThrValGluA 948

     3254 AACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCC 3303
          ::   |||:::   |||   :::   :::      :::|||   |||:::
      948 spAlaAlaLysIleAspAlaThrGlyGlyAlaAlaThrLeuThrAlaThr 964

     3304 GGGCGCAATGCCACCGAAAAGGCAGAAAGT 3333
          :::   :::   |||   ||||||:::|||
      965 SerGlyLysLeuThrThrLysAlaSerSer 974
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41725

seq_documentation_block:

```
ID   AAR41725 standard; Protein; 1536 AA.
XX
AC   AAR41725;
XX
DT   26-APR-1994  (first entry)
XX
DE   High molecular weight protein 1 (HMW1).
XX
KW   HMW; high molecular weight protein; virus; vaccine; influenza;
KW   epitope; immunity; haemophilus influenzae; gene cluster.
XX
OS   Haemophilus influenzae.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 668..677
FT                   /note= "Possibly incorrect sequence. Alternative
FT                           sequence for this region is LNVSESGEFN.
FT                           (See comments)."
XX
PN   WO9319090-A.
XX
PD   30-SEP-1993.
XX
PF   16-MAR-1993;   93WO-US02166.
XX
PR   16-MAR-1992;   92GB-0005704.
XX
PA   (BARE/) BARENKAMP S J.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1993-320683/40.
DR   N-PSDB; AAQ49508.
XX
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties
XX
PS   Claim 3; Figure 2/10; 100pp; English.
XX
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.  This sequence is claimed to be
CC   the same as that given in AAR41723 (High molecular weight protein 1)
CC   although it does differ slightly. (Repeated regions which are
CC   possibly incorrect and occur in the corresponding nucleotide coding
CC   sequence contribute to these differences).
XX
SQ   Sequence   1536 AA;
```

alignment_scores:

```
            Quality:  266.00              Length:  1297
              Ratio:  0.439                 Gaps:    73
 Percent Similarity:  46.723   Percent Identity: 20.971
```

```
alignment_block:
 US-09-303-518D-653 x AAR41725   ..

 Align seg 1/1  to: AAR41725  from: 1  to: 1536

    178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
        |||   |||   :::::|||||||   ||||||   :::|||
    199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal........ 211

    228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
              :::|||:::|||   ::::::   :::
    212 .......AsnLeuIleGlyGlyLysValLysAsnGlu............. 221

    278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
          :::|||:::|||   ||||||      ::::::   ||||||||||:::
    222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

    322 CAA.....................................TATATTGTGAG 335
        :::                                    |||   :::::
    238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

    336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
        ::::   :::|||   :::   |||      |||:::   |||:::|||:
    254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271

    386 GCAAT.......................CCCGATCAGCACCGCTTTTCT 411
        ::|||                       :::|||   :::::::|||
    271 lyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287

    412 TACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCC 461
          :::   |||:::::::::      |||:::|||
    288 AlaAspSerValSerLysAsp......LysSerGly.............. 297

    462 TTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATG 511
                                       :::::::|||   :::|
    298 ................................AsnIleValLeuSerA 303

    512 CAGAACCTGTTGAG.................ATGACCAGTTATATGGAT 543
        ||:::      |||                 :::::::::      :::
    303 laLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319

    544 GGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT......CG 587
              |||   :::   |||            |||:::|||      ::
    320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLy 336

    588 AATCGGAGCA.........................GGCAGACAATATT 610
        :   ||||||                        |||:::   |||
    336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353

    611 GGCGGTCTGATGAA...GACGAACCCAATAAC.................... 639
              :::||||||      |||   :::|||
    353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369

    640 ..............CGCGAAAGTTCATATCATATTGCAAGC........ 666
                        :::   |||:::   ::::::::::::
    370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLy 386

    667 ..........GCATATTCTTGG............CTCGTCGGTGGCAATA 694
                |||      |||             |||:::   ||||||
    386 sGlyGlyArgAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnI 403

    695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAA 744
              ||||||   |||||||||
    403 leAsnAlaGln...GlySerGly........................... 409

    745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
                                  :::      |||||||||
    410 ....................AspIleAlaLysThrGlyGlyPheValGl 419

    795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGT 844
            ||||||      :::||||||   |||
    419 uThrSerGlyHisAspLeuPheIleLysAsp................... 429

    845 TAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAAT 894
            |||::::::::::::::::
    430 .....AsnAlaIleValAspAla........................... 435

    895 GGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGG 944
                          |||:::|||:::   |||
    436 ..................LysGluTrpLeuLeuAspPro........... 442

    945 AGATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTT 994
        |||:::   |||:::      |||         :::::::::
    443 .AspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerG 459

    995 TTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACAC 1044
          :::|||:::   :::|||:::|||                    :::
    459 luAspAspGluTyrThrGlySerGly..................AsnSer 469

   1045 TATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTTAATGT 1094
          |||   |||   |||   |||   :::|||:::   |||   |||::
    470 AlaSerThrProLysArgAsnLysGluLysThrThr...LeuThrAsnTh 485

   1095 TTCTTTATCCGAG.............................ACAGCAA 1114
        ::::|||::::::                             |||||||:
    485 rThrLeuGluSerIleLeuLysLysGlyThrPheValAsnIleThrAlaA 502

   1115 GAGAACCTGTTTAT..................................... 1128
        :::::   :::|||
    502 snGlnArgIleTyrValAsnSerSerIleAsnLeuSerAsnGlySerLeu 518

   1129 ..................CATGCTGCAGGTGGGGTCAACAGTTATCGACC 1160
                           :::::::::|||||||||
    519 ThrLeuTrpSerGluGlyArgSerGlyGlyGlyVal.............. 530

   1161 CAGACTGAATAAT..........................GGAGAAAATA 1183
        :::::|||||||                          |||   |||:
    531 .GluIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsnL 547

   1184 TTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATC 1233
        ::::::::   :::   |||      :::      :::   :::|||||||
    547 euThrIleTyrSerGlyGlyTrpValAsp......ValHisLysAsnIle 561

   1234 AACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCC 1283
        :::   ||||||                 :::||||||:::::::::::::
    562 SerLeuGlyAla..............GlnGlyAsnIleAsnIleThrAl 573

   1284 TAAA........................AACAACGAAACGTGGCAAGGCG 1309
          |||                       :::|||::::::      |||
    573 aLysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyG 590

   1310 CGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGC 1359
          |||:::   |||:::::::|||:::         ::::::   |||:::
    590 lnGlyThr...IleThrSerGlyAsnGlnLysGlyPheArgPheAsnAsn 605

   1360 GTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCA 1409
        |||:::         |||:::      |||:::|||   |||
    606 ValSer.........LeuAsnGlyThrGlySerGly...LeuGlnPheTh 618

   1410 AGCCAAAGGGGAAAAC.......................CAAGGCTCGG 1435
          :::|||      |||                      :::|||::::
    618 rThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyThrL 635

   1436 TCAGCGTGGGCGACGGTAAAGTCATCTTA..................GAT 1467
        ::::::::   :::|||||||||   :::                  :::
    635 euAsnIle...SerGlyLysValAsnIleSerMetValLeuProLysAsn 650

   1468 CAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTT............AG 1505
        :::::::::   |||:::   |||   :::::::::            ||
    651 GluSerGlyTyrAspLysPheLysGlyArgThrTyrTrpAsnLeuThrSe 667

   1506 TGAAATCGGCTTGGTCAGCGGCAGG.......................... 1530
        |::::::   ::::::::::::::
    667 rLysValAspMetIleAsnSerLysAspAlaLeuThrIleAspSerArgG 684
```

```
    1243 GluValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLy 1259

    2949 GCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTC.......... 2988
         |   |||   ::::::|||::::::   |||   |||:::
    1259 s...AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleG 1275

    2989 ..............AACAATACCGGCAACGAACCCGTAAGTCTCGAGCAA 3024
                       :::||||||   |||      ::::::      :::
    1275 lyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1291

    3025 TTGACGGTA.......................GTGGAAGGAAAAGACAA 3050
         |||||||||                       :::||||||      ::
    1292 LeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaTh 1308

    3051 CACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCG 3100
         :         |||::::::   |||      ::::::|||:::
    1308 rLeuThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleT 1325

    3101 ATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGC 3150
           ::: |||   |||   ::::::||||||||||:::   :::|||   ||
    1325 hrSer.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAl 1341

    3151 CTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACA 3187
         |      |||::::::   |||   ::::::   |||
    1341 aGlySerIleAsnAlaAlaAsnValThrLeuAsnThr 1353

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37120

seq_documentation_block:
ID   AAU37120 standard; Protein; 2344 AA.
XX
AC   AAU37120;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #1290.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS54979.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 12713; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2344 AA;


alignment_scores:
           Quality:  266.00              Length:   1581
             Ratio:  0.358                 Gaps:     63
 Percent Similarity:  47.059   Percent Identity: 17.710

alignment_block:
 US-09-303-518D-653 x AAU37120   ..

 Align seg 1/1  to: AAU37120  from: 1  to: 2344

    163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATAT 212
        |||   :::   ::::::|||:::             |||   |||:::||
     20 ArgLeuTyrLysSerGlyLysAsnTrpValLysSerGlyIleLysGluIl 36

    213 TGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAG 262
        ||||:::::::::      |||
     36 eGluMetPheLysIleMetGly.........................L 44

    263 CCCCGATGATTGATTTTTCTGTGGTATCGCGTAACGGC............ 300
          |||:::|||:::   |||:::||||||:::::::::
     44 euProPheIleSerHisSerMetValSerGlnAspAsnGlnSerIleSer 60

    301 ........................GTGGCGGCATTGGCGGGCGATCA 323
                                :::::|||:::   |||
     61 LysLysMetThrGlyTyrGlyLeuLysThrThrAlaValIleGlyGlyAl 77

    324 ATATATTGTGAGCGTGGCACATAAC........................ 348
         :::   |||::::::   |||:::
     77 aPheThrValAsnMetLeuHisAspGlnGlnAlaPheAlaAlaSerAspA 94

    349 ....................................GGCGGCTAT 357
                                            |||:::
     94 laProLeuThrSerGluLeuAsnThrGlnSerGluThrValGlyAsnGln 110

    358 AACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCGAT........... 396
        |||::::::   :::   |||:::   |||:::   |||
    111 AsnSerThrThrIleAspAlaSerThrSerThrAlaAspSerThrSerVa 127

    397 ....CAGCACCGCTTTTCTTACCAAATT..................... 420
            ::::::      |||   |||
    127 lThrLysAsnSerSerSerValGlnThrSerAsnSerAspThrValSerS 144

    421 ..GTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCCTTATGGC 468
            |||   :::|||      :::   ||||||:::
    144 erGluLysSerGluAsnValThrSerThrThrAsnSer............ 156

    469 GGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATGCAGAACC 518
                                              |||:::   :::
    157 ................................ThrSerAsnGlnGl 161

    519 TGTTGAGATGACCAGTTATATGGATGGGTGGAAATACGCTGATTTAAATA 568
        ::::::||||||     ::::::   :::      :::   ::::
    161 nGluLysLeuThrSerThrSerGluSerThrSerSerLysAsnThrThrS 178
```

```
     3826 GACGGCATCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCA 3875
          |:::   |||   :::                 ::::::     |||   :
     1153 rSerMetSerAspSer.............ThrSerLeuSerThrSerGluS 1166

     3876 ATACGGCATCGGCAGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTA 3925
          ::   :::::::::   |||||||||:::|||
     1166 erAspSerThrSerAspSerThrSerThrSerAspSerIleSerGluAla 1182

     3926 GTAGCGGCAGCCTTTCAGACGGCA......................... 3949
          ::::::::::::   :::|||:::
     1183 IleSerGlySerGluSerThrSerIleSerLeuSerGluSerAsnSerTh 1199

     3950 .......TCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCA 3992
                  ||||||:::||||||||||:::|||           ::::::
     1199 rSerAspSerGluSerLysSerAlaSerAlaPheLeuSerGluSerLeuS 1216

     3993 GGCAAGATACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCG 4042
            :::::::|||:::::::   :::::::::|||   |||:::   |||||
     1216 erGluSerThrSerGluSerThrSerGluSerLeuSerGlySerThrSer 1232

     4043 GCGCAACGCGCTATTTCG...TCCAAAAAGCGGATTACCGATACGAAAAC 4089
             :::   ::::::|||   |||:::::::::      :::|||:::||
     1233 AspSerThrSerLeuSerAspSerAsnSerGluSerGlySerThrSerTh 1249

     4090 GTCAATATCG.........CCACCCCGGGCCTTGCATTCAACCGCTACCG 4130
          ||||:::|||                  |||      ||||||:::|||:
     1249 rSerLeuSerAsnSerThrSerGlySerAlaSerIleSerThrSerThrS 1266

     4131 CGCG..............................GGCATTAAGGCAGAT 4149
          ::                                ::::::::::::::
     1266 erGlySerAlaSerThrSerThrValLysSerGluSerValSerThrSer 1282

     4150 TATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCT 4199
             |||   :::   :::      :::|||   :::      ||||||
     1283 LeuSerThrSerThrSerThrSerLeuSerAspSerThrSerLeuSerTh 1299

     4200 GTCCTATACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCG 4249
           |||   :::|||::::::||||||:::      :::   ::::::::::
     1299 rSerLeuSerAspSerThrSerGlySerLysSerAsnSerLeuSerAlaS 1316

     4250 CCGTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAA 4290
          :::::   ::::::|||      :::||||||::::::|||
     1316 erMetSerThrSerAspSerIleSerThrArgLysSerGlu 1329
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01835

seq_documentation_block:

ID AAB01835 standard; Protein; 1095 AA.
XX
AC AAB01835;
XX
DT 11-SEP-2000 (first entry)
XX
DE Haemophilus influenzae strain LCDC2 mature HMW1A protein, SEQ ID NO:45.
XX
KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
OS Haemophilus influenzae strain LCDC2.
XX
PN WO200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
DR WPI; 2000-303789/26.
DR N-PSDB; AAA52184.
XX
PT Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
PS Claim 8; Fig 22A-P; 307pp; English.
XX
CC The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMWA
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typeable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents a mature HMWA protein from a non-typeable
CC strain of H. influenzae.
XX
SQ Sequence 1095 AA;

alignment_scores:

```
            Quality:  265.50              Length:   1042
              Ratio:  0.503                 Gaps:     55
 Percent Similarity:  50.672   Percent Identity: 21.881
```

alignment_block:

US-09-303-518D-653 x AAB01835 ..

Align seg 1/1 to: AAB01835 from: 1 to: 1095

```
      574 CCTGATCGTGTTCGAATCGGAGCAGGCAGACAATATTGGCGGTCTGATGA 623
          ||||||:::|||   ||||||||||||   :::      |||||||||::
        1 ProAspGluValThrIleGlyAlaGly...AspValGlyArgSerAspAs 16

      624 AGAC..............................GAACCCAATAACCGCG 643
          ::::                              |||   |||:::   :
       16 pSerSerAspThrAlaPheProThrGlyThrGlyGluArgAsnSerProL 33

      644 AAAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAAT 693
          ::::::::   :::          :::   :::                |||
       33 ysThrAsnAlaGlnAsnArgProThrIleThr..............Asn 44

      694 ACCTTTGCACAAAATGGATCAGGTGGTGGCACA...GTCAACTTAGGTAG 740
          |||      ::::::         :::|||||   |||||||:::   ::
       45 ThrSerLeuGluGlnIleLeuLysAsnGlyThrPheValAsnIleThrAl 61

      741 CGAA..........................AAAATTAAACATAGCCCAT 763
          ::::                          :::||||||::::::   :
       61 aLysAsnLysIleLeuValAsnSerAspIleAsnIleLysGluAsnSerH 78

      764 AT..............................GGTTTTTTACCAACA 780
```

```
                ::                                          |||
  78 isLeuIleLeuTrpSerGluArgAspGlyAsnSerGlyValGlnIleAsp 94

 781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTAT........ 822
     |||:::       :::   :::|||:::   :::   :::|||
  95 GlyAsnIleThrSerAlaThrGlyGlySerLeuThrValTyrSerSerGl 111

 823 .......GATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTG.... 861
            |||:::::|||:::      |||    :::|||   |||
 111 yTrpValAspValHisLysAsnIleThrLeuAsnSerGlyTyrLeuAsnI 128

 862 ........CAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAG 903
             ::::::|||:::      :::   ::::::|||   :::
 128 leThrThrLysSerGlyAspValAlaPheGluGlnGlyAsnAspLeuThr 144

 904 CTAGTT.......................CGTAAAGATTGGTTCTATGA 929
     ::::::                        :::|||   :::   :::||
 145 IleThrGlyGlnGlyThrIleThrAlaSerLysLysGlyPheArgPheAs 161

 930 TGAAATCTTTGCTGGAGATACCCATTCAGTATTC.........TACGAAC 970
     |::::::       :::   :::   :::   |||        |||:::
 161 pAsnValThrLeuSerGlyValLysLysGlyPheLeuPheLysTyrSerG 178

 971 CACATCAAAATGGGAAA...............TACTTTTTTAACGACAAT 1005
         :::|||:::|||                   |||:::   :::
 178 lnThrAsnAsnAsnLysAspSerAsnPheGluAsnHisPheArgGlyThr 194

1006 AATAATGGCGCAGGAAAAATCGAT...........GCCAAACATAAACA 1043
        |||   :::|||||||:::|||          |||::::::::::::
 195 LeuAsnIleSerGlyLysValAspIleLeuMetGlnAlaArgGlnGluAs 211

1044 CTATTCTCTACCTTAT..........................AGATTAA 1066
     ::::::      :::                            ||||||:
 211 nTrpAsnArgArgHisSerGlyArgSerHisTrpAsnValThrArgLeuA 228

1067 AAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGA 1116
     :::::   |||   :::   :::|||::::::::::::::::::::
 228 snValSerThrAsnSerTyrLeuAsnIleThrIleAspAsnSerGlySer 244

1117 GAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACT 1166
     :::|||        :::   |||:::          ||||||
 245 ArgPro.........SerProGlyAlaGlyProLeuTyrArgArgSerGl 258

1167 GAATAATGGA.........GAAAATATTTCCTTTATTGACAAAGGAAAAG 1207
        ||||||            :::|||   :::   ::::::   ::::::|
 258 yLeuAsnGlyIleSerPheAsnAsnAspThrValPheAsnValAlaSerG 275

1208 GTGAATTGATA....................CTTACCAGCAACATCAAC 1236
     ||:::   :::                         ::::::||||||::::::
 275 lySerAlaValAsnPheSerIleLysProProIleValSerAsnValHis 291

1237 CAAGGCGCGGGCGGTTTGTATTTTGAGGGTAAT................ 1269
     :::|||          |||   |||:::||||||
 292 AspGlyAsnHisThrLeu...PheAsnGlyAsnValSerValLeuGlyGl 307

1270 ..................TTTACGGTCTCGCCTAAAAACAACGAAACGT 1300
                       |||::::::|||   :::|||:::   |||
 307 yGlyAspValAsnPheHisPheAsnAlaSerSerSerAsnHisTrpThrH 324

1301 GG...........CAAGGCGCGGGCGTTCATATCAGTGATGGCAGTACC 1338
                  ::::::   :::   :::   |||:::||||||:::
 324 isGlyValValIleLysSerGlnAsnPheAsnAlaSerGluGlySerSer 340

1339 GTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGG 1388
     :::   :::|||   :::|||      :::   |||   :::
 341 LeuArgPheLysSerGluGly......SerThrArgThrAlaPheThrIl 355

1389 CAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCA 1438
      ::::::   |||   ::::::|||   |||
 355 eGluSerAspLeuThrLeuAsnAlaThrGly.................. 365

1439 GCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGC 1488
                 |||::::::   |||:::|||   |||      :::|||
 366 ..........GlyAsnIleSerLeuAsnGlnValAlaGlyIleAspGly 378

1489 AAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGG........ 1530
     :::   |||                :::||||||:::::::::
 379 AsnLeuGlnLys............SerLeuValAlaAsnLysAsnIleTh 391

1531 ..........GGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCG 1570
                |||::::::   |||   |||                  |
 391 rPheGluGlyGlyAsnIleThrLeuAlaAla..................A 402

1571 ACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCAT 1620
     |||||         :::   ::::::|||:::          :::|||
 402 spLysLysProIleGluIleLysGlyAsnIleThrValLysGluGlyAla 418

1621 TCGCTTTCGTTCCACCGCATTCAA...AATACCGATGAAGGGGCGATGAT 1667
     :::::::::::::::      :::   ::::::|||::::::|||:::
 419 AsnValThrLeuArgSerAlaAsnTyrGlyAsnAspLysSerAlaLeuSe 435

1668 TGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAAT. 1716
      ::::::   |||      ::::::::::::::|||:::|||||||:::
 435 rIleArgGlyAsnValThrAsnLysGlyAsnLeuThrValThrGlySerA 452

1717 ..............AAAGATATTACTACAACCGGCAAT............ 1740
                   |||::::::|||:::   |||:::
 452 laIleAsnIleGluLysAsnLeuThrValGluGlySerAlaLysPheLeu 468

1741 .....................................AACAACAACTT 1751
                                          :::|||:::
 469 AlaAsnProAsnTyrSerPheAsnValSerGlyLeuPheAspAsnGlnGl 485

1752 GGATAGCAAAAAAGAAATTGCCTACAACGGT...TGGTTTGGCGAGAAAG 1798
         |||:::   :::||||||   :::|||      |||   :::   :
 485 yLysSerAsnIleSerIleAlaLysGlyGlyAlaHisPheLysAspIleA 502

1799 ATGCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAAGAA 1848
     ::   ||||||:::            ||||||::::::   :::   ::::::
 502 snAsnThrLysSer.........LeuAsnIleThrThrAsnSerAspSer 515

1849 GCGGATCGCACTTTACTGCTTTCCGGCGGAACAAATTTAAACGGCAATAT 1898
     |||   ||||||::::::   ::::::   ||||||   ||||||:::::
 516 AlaTyrArgThrIleIleGluGlyAsnIleThrAsnSerAsnGlyAspLe 532

1899 CACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACG 1948
     ::::   |||::::::|||                           :::|
 532 uAsnIleThrAspAsnLys.......................AsnAsnA 541

1949 CCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAA 1998
     ||        ::::::|||::::::   |||:::   ||||||
 541 laGluIleGlnIleGlyGlyAsnIleSerGlnLysGluGly......... 554

1999 GGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGA 2048
                                   :::   |||::::::::::::
 555 .............................AsnLeuThrIleSerSerAs 561

2049 AAACTTCCATATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCA 2098
     ::::::::::|||      :::|||   ::::::::::::::|||   |
 561 pLysIleAsnIleThr...AsnGlnIleThrIleLysLysGlyValAsnL 577

2099 AAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGT 2148
     ||   ::::::|||          :::|||:::|||:::
 577 ysGluAspSerAspSerSerThrAlaAsnAsnAlaAsnLeu......... 590

2149 GTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGG 2198
                          ||||||   |||:::
 591 .....................ThrIleLysThrLys............. 595

2199 TCTGACAAGTTGTACCGAAAAAACCATTACCGACGATAAAGTGATTGCTT 2248
                        |||      :::|||   |||      |||:::
 596 ................GluLeuGlnLeuThrGlyAspLeuAsnIleSerG 607

2249 CATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGATCACGCT 2298
     :::::::|||::::::|||             :::      :::   |||
 607 lyPheAspLysAlaGluIle...........ThrAlaLysGluGlyAla 619
```

```
2299 CATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGCAGG 2348
620 AspLeu...IleIleGlyAsnSerAspAsnAsnAsnAsnAlaAsnAla.. 634

2349 CGGAGACACGCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACC 2398
635 .............LysLysValThrPheAsnGlnValLysAspSerLysI 647

2399 TCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAA......GCCACA 2442
647 leSer...AlaAspSerHisAsnValThrLeuAsnSerLysValGluThr 662

2443 TTAAACGGCAACACATCGGCTTCGGACAAT.................GC 2474
663 SerAsnGlyAsnAsnAspAlaGluSerAsnAsnGlyAspGlyThrSerLe 679

2475 TTCATTTAATCTAAGCAACAACGCCGTACAAAACGGC.............. 2511
679 uThrIleAsnAlaLysAsnIleThrValAsnAsnAsnIleThrSerHisL 696

2512 ........AGTCTGACGCTTTCCGACAAC......GCTAAGGCAAACGTA 2547
696 ysThrValAsnIleThrAlaSerGluAsnValThrThrLysAlaGlyThr 712

2548 AGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATT 2597
713 ThrIleAsnAlaThrThrGlySerValGluValThrAlaLys........ 726

2598 CCATTTTGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATA 2647
727 .....ThrGlyAspIleLysGlyLysValGluS 736

2648 CGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAA 2697
736 erThr...........SerGlySerValThrLeuThrAlaThrGlyGlu 748

2698 .........TTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAA 2738
749 AlaLeuAlaValSerAsnIleSerGlyAsnThrValThrIleThrAlaAs 765

2739 TTCCGCCTATCGACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAG 2788
765 nLysGlyLysLeuThrThrGlnAlaGlySer...ThrValSerAlaIleA 781

2789 ATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCA 2838
781 snGlyValThrAlaSerSerGlnSerGlyAspIleSerGlyThr..... 795

2839 ACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAA 2888
796 .................IleSerGlyAsnThrValLysValSerAlaIl 806

2889 CGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCG 2938
806 eGlyAspLeuThrThrLysSerGlySerGluIle.............. 817

2939 GCAAATTGAAGCTGGCGGAA.............AGTTCCGAAGGCACTTAC 2976
818 ..LysAlaLysThrGlyGluAlaAsnValThrSerAlaThrGlyThrIle 833

2977 ......ACCTTG.........GCTGTCAAC.........AATACCGGCAA 3002
834 GlyGlyThrIleSerGlyAsnAlaValAsnValThrAlaAsnThrGlyAs 850

3003 CGAACCCGTA............................AGTCTCGAGC 3022
850 pLeuThrValGluAspAlaAlaLysIleAspAlaThrGlyGlyAlaAlaT 867

3023 AATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTT 3072
867 hrLeuThrAlaThrSerGlyLysLeuThrThrLysAlaSerSerSerIle 883

3073 AATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCA 3122
884 Thr.Ser.........................AlaAsnAsnGlnValA 891

3123 GCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAAC 3172
891 snLeuSerAlaLysAspGlySerIleGlyGlyAsnIleAsnAlaAlaAsn 907

3173 AAGAGCTTTCCGACAAACTCGGCA 3196
908 ValThrLeuAsnThrThrGlyAla 915
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01834

seq_documentation_block:

```
ID   AAB01834 standard; Protein; 1101 AA.
XX
AC   AAB01834;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain LCDC2 HMW1A protein, SEQ ID NO:43.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain LCDC2.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52183.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 22A-P; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
```

```
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1101 AA;


alignment_scores:
            Quality:  265.50              Length:   1042
              Ratio:  0.503                 Gaps:     55
 Percent Similarity:  50.672   Percent Identity: 21.881

alignment_block:
 US-09-303-518D-653 x AAB01834   ..

 Align seq 1/1  to: AAB01834  from: 1  to: 1101

     574 CCTGATCGTGTTCGAATCGGAGCAGGCAGACAATATTGGCGGTCTGATGA 623
         ||||||:::|||   ||||||||||||   :::      |||||||||::
       7 ProAspGluValThrIleGlyAlaGly...AspValGlyArgSerAspAs 22

     624 AGAC............................GAACCCAATAACCGCG 643
         ::::                           |||   |||:::   :
      22 pSerSerAspThrAlaProThrGlyThrGlyGluArgAsnSerProL 39

     644 AAAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAAT 693
         ::::::::   :::         :::   :::                |||
      39 ysThrAsnAlaGlnAsnArgProThrIleThr..............Asn 50

     694 ACCTTTGCACAAAATGGATCAGGTGGTGGCACA...GTCAACTTAGGTAG 740
         |||      ::::::      :::||||||   ||||||:::   ::
      51 ThrSerLeuGluGlnIleLeuLysAsnGlyThrPheValAsnIleThrAl 67

     741 CGAA..........................AAAATTAAACATAGCCCAT 763
         ::::                         :::||||||::::::   :
      67 aLysAsnLysIleLeuValAsnSerAspIleAsnIleLysGluAsnSerH 84

     764 AT..............................GGTTTTTTACCAACA 780
         ::                           |||
      84 isLeuIleLeuTrpSerGluArgAspGlyAsnSerGlyValGlnIleAsp 100

     781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTAT........ 822
         |||:::      :::   :::|||:::   :::   :::|||
     101 GlyAsnIleThrSerAlaThrGlyGlySerLeuThrValTyrSerSerGl 117

     823 .......GATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTG.... 861
             |||:::::::|||:::      |||   :::|||   |||
     117 yTrpValAspValHisLysAsnIleThrLeuAsnSerGlyTyrLeuAsnI 134

     862 ........CAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAG 903
                 ::::::|||:::      :::   ::::::|||   :::
     134 leThrThrLysSerGlyAspValAlaPheGluGlnGlyAsnAspLeuThr 150

     904 CTAGTT........................CGTAAAGATTGGTTCTATGA 929
         ::::::                       :::|||   :::   :::||
     151 IleThrGlyGlnGlyThrIleThrAlaSerLysLysGlyPheArgPheAs 167

     930 TGAAATCTTTGCTGGAGATACCCATTCAGTATTC.........TACGAAC 970
         |::::::      :::   :::   :::   |||         |||:::
     167 pAsnValThrLeuSerGlyValLysLysGlyPheLeuPheLysTyrSerG 184

     971 CACATCAAAATGGGAAA..............TACTTTTTTAACGACAAT 1005
               :::|||:::|||                        |||:::   :::
     184 lnThrAsnAsnAsnLysAspSerAsnPheGluAsnHisPheArgGlyThr 200

    1006 AATAATGGCGCAGGAAAAATCGAT............GCCAAACATAAACA 1043
            |||   :::|||||||:::|||            |||:::::::::::
     201 LeuAsnIleSerGlyLysValAspIleLeuMetGlnAlaArgGlnGluAs 217

    1044 CTATTCTCTACCTTAT........................AGATTAA 1066

         :::::::      :::                              ||||||:
     217 nTrpAsnArgArgHisSerGlyArgSerHisTrpAsnValThrArgLeuA 234

    1067 AAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGA 1116
         :::::   |||   :::   :::|||:::::::::::::::::::::::
     234 snValSerThrAsnSerTyrLeuAsnIleThrIleAspAsnSerGlySer 250

    1117 GAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACT 1166
         :::|||         :::   |||:::         ||||||
     251 ArgPro.........SerProGlyAlaGlyProLeuTyrArgArgSerGl 264

    1167 GAATAATGGA.........GAAAATATTTCCTTTATTGACAAAGGAAAAG 1207
             ||||||         :::|||   :::   ::::::   :::::::|
     264 yLeuAsnGlyIleSerPheAsnAsnAspThrValPheAsnValAlaSerG 281

    1208 GTGAATTGATA.....................CTTACCAGCAACATCAAC 1236
         ||:::   :::                     ::::::||||||::::::
     281 lySerAlaValAsnPheSerIleLysProProIleValSerAsnValHis 297

    1237 CAAGGCGCGGGCGGTTTGTATTTTGAGGGTAAT.................. 1269
         :::|||         |||   |||:::||||||
     298 AspGlyAsnHisThrLeu...PheAsnGlyAsnValSerValLeuGlyGl 313

    1270 ...................TTTACGGTCTCGCCTAAAAACAACGAAACGT 1300
                             |||::::::|||   :::|||:::   |||
     313 yGlyAspValAsnPheHisPheAsnAlaSerSerSerAsnHisTrpThrH 330

    1301 GG............CAAGGCGCGGGCGTTCATATCAGTGATGGCAGTACC 1338
                      ::::::   :::   :::   |||:::||||||:::
     330 isGlyValValIleLysSerGlnAsnPheAsnAlaSerGluGlySerSer 346

    1339 GTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGG 1388
         :::   :::|||   :::|||      :::   |||   :::
     347 LeuArgPheLysSerGluGly......SerThrArgThrAlaPheThrIl 361

    1389 CAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCA 1438
          ::::::   |||   ::::::|||   |||
     361 eGluSerAspLeuThrLeuAsnAlaThrGly..................... 371

    1439 GCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGC 1488
                  |||::::::   |||:::|||   |||      :::|||
     372 ...........GlyAsnIleSerLeuAsnGlnValAlaGlyIleAspGly 384

    1489 AAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGG........ 1530
         :::   |||                ::::::::::::::::
     385 AsnLeuGlnLys............SerLeuValAlaAsnLysAsnIleTh 397

    1531 ..........GGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCG 1570
                    |||::::::   |||   |||                 |
     397 rPheGluGlyGlyAsnIleThrLeuAlaAla..................A 408

    1571 ACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCAT 1620
         |||||      :::   ::::::|||:::            :::|||
     408 spLysLysProIleGluIleLysGlyAsnIleThrValLysGluGlyAla 424

    1621 TCGCTTTCGTTCCACCGCATTCAA...AATACCGATGAAGGGGCGATGAT 1667
         ::::::::::::::      :::   ::::::|||::::::|||:::
     425 AsnValThrLeuArgSerAlaAsnTyrGlyAsnAspLysSerAlaLeuSe 441

    1668 TGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAAT. 1716
          ::::::   |||      :::::::::::::::|||:::|||||||:::
     441 rIleArgGlyAsnValThrAsnLysGlyAsnLeuThrValThrGlySerA 458

    1717 ..............AAAGATATTACTACAACCGGCAAT............ 1740
                       |||::::::|||:::   |||:::
     458 laIleAsnIleGluLysAsnLeuThrValGluGlySerAlaLysPheLeu 474

    1741 ....................................AACAACAACTT 1751
                                             :::|||:::
     475 AlaAsnProAsnTyrSerPheAsnValSerGlyLeuPheAspAsnGlnGl 491

    1752 GGATAGCAAAAAAGAAATTGCCTACAACGGT...TGGTTTGGCGAGAAAG 1798
         |||:::   :::|||||||   :::|||      |||   :::   :
```

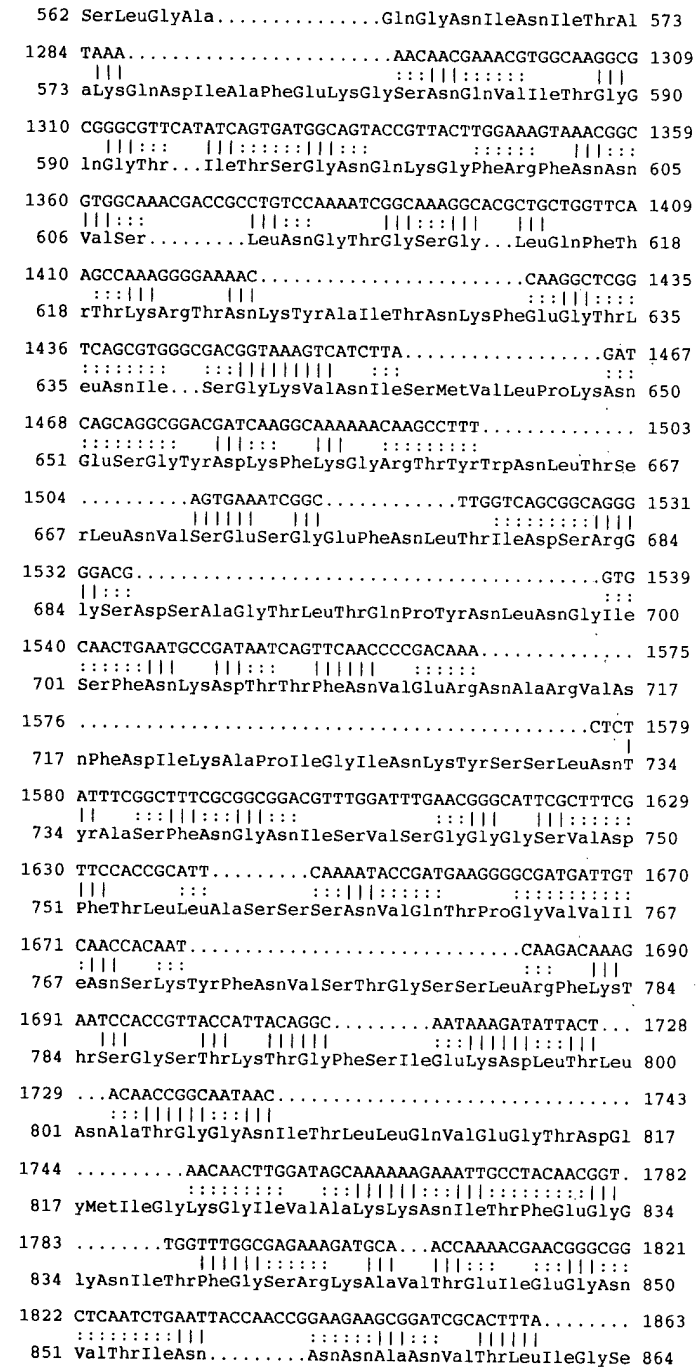
```
 491 yLysSerAsnIleSerIleAlaLysGlyGlyAlaHisPheLysAspIleA 508

1799 ATGCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAAGAA 1848
     ::   ||||||:::         ||||||:::::::   :::   ::::::
 508 snAsnThrLysSer.........LeuAsnIleThrThrAsnSerAspSer 521

1849 GCGGATCGCACTTTACTGCTTTCCGGCGGAACAAATTTAAACGGCAATAT 1898
     |||   ||||||::::::   ::::::   ||||||   ||||||:::::
 522 AlaTyrArgThrIleIleGluGlyAsnIleThrAsnSerAsnGlyAspLe 538

1899 CACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACG 1948
     ::::   |||:::::::|||                          :::|
 538 uAsnIleThrAspAsnLys.......................AsnAsnA 547

1949 CCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAA 1998
     ||      ::::::|||::::::   |||:::   ||||||
 547 laGluIleGlnIleGlyGlyAsnIleSerGlnLysGluGly......... 560

1999 GGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGA 2048
                                  :::   |||:::::::::::
 561 ............................AsnLeuThrIleSerSerAs 567

2049 AAACTTCCATATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCA 2098
     ::::::::::|||      :::|||   ::::::::::::::|||   |
 567 pLysIleAsnIleThr...AsnGlnIleThrIleLysLysGlyValAsnL 583

2099 AAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGT 2148
     ||   :::::::|||         :::|||:::|||:::
 583 ysGluAspSerAspSerSerThrAlaAsnAsnAlaAsnLeu......... 596

2149 GTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGG 2198
                           ||||||   |||:::
 597 .....................ThrIleLysThrLys.............. 601

2199 TCTGACAAGTTGTACCGAAAAAACCATTACCGACGATAAAGTGATTGCTT 2248
                    |||      :::|||   |||      |||::::
 602 ................GluLeuGlnLeuThrGlyAspLeuAsnIleSerG 613

2249 CATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGATCACGCT 2298
     ::::::::|||:::::::|||           :::      :::   |||
 613 lyPheAspLysAlaGluIle............ThrAlaLysGluGlyAla 625

2299 CATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGCAGG 2348
        |||   :::   |||   :::      |||:::|||   :::|||
 626 AspLeu...IleIleGlyAsnSerAspAsnAsnAsnAsnAlaAsnAla.. 640

2349 CGGAGACACGCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACC 2398
                       ||||||   |||   ::::::::::::::::
 641 .............LysLysValThrPheAsnGlnValLysAspSerLysI 653

2399 TCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAA......GCCACA 2442
     ::|||   :::   :::   :::::::|||:::|||:::          |||
 653 leSer...AlaAspSerHisAsnValThrLeuAsnSerLysValGluThr 668

2443 TTAAACGGCAACACATCGGCTTCGGACAAT.................GC 2474
        |||||||||:::::|||:::::|||
 669 SerAsnGlyAsnAsnAspAlaGluSerAsnAsnGlyAspGlyThrSerLe 685

2475 TTCATTTAATCTAAGCAACAACGCCGTACAAAACGGC.............. 2511
      :::::::|||   :::|||   :::|||:::|||:::
 685 uThrIleAsnAlaLysAsnIleThrValAsnAsnAsnIleThrSerHisL 702

2512 ........AGTCTGACGCTTTCCGACAAC......GCTAAGGCAAACGTA 2547
             :::::::|||   |||:::|||      :::||||||::::::
 702 ysThrValAsnIleThrAlaSerGluAsnValThrThrLysAlaGlyThr 718

2548 AGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATT 2597
     :::   :::|||   :::|||:::|||::::::::::   |||
 719 ThrIleAsnAlaThrThrGlySerValGluValThrAlaLys........ 732

2598 CCATTTTGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATA 2647
                           ||||||   |||:::|||      ::::
 733 .....................ThrGlyAspIleLysGlyLysValGluS 742

2648 CGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAA 2697
     :::::            :::::::::   ||||||   :::      |||
 742 erThr............SerGlySerValThrLeuThrAlaThrGlyGlu 754

2698 .........TTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAA 2738
                ::::::|||::::::   :::::::::|||||||||   ||
 755 AlaLeuAlaValSerAsnIleSerGlyAsnThrValThrIleThrAlaAs 771

2739 TTCCGCCTATCGACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAG 2788
     |::::::                 ||||||:::   |||   ||||||   :
 771 nLysGlyLysLeuThrThrGlnAlaGlySer...ThrValSerAlaIleA 787

2789 ATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCA 2838
     :::::              |||:::   :::   :::|||   |||
 787 snGlyValThrAlaSerSerGlnSerGlyAspIleSerGlyThr...... 801

2839 ACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAA 2888
                         ::::::      ||||||:::   :::
 802 ..................IleSerGlyAsnThrValLysValSerAlaIl 812

2889 CGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCG 2938
      |||:::   |||   :::      ||||||:::
 812 eGlyAspLeuThrThrLysSerGlySerGluIle.............. 823

2939 GCAAATTGAAGCTGGCGGAA............AGTTCCGAAGGCACTTAC 2976
       |||   |||   :::|||              |||:::   ||||||
 824 ..LysAlaLysThrGlyGluAlaAsnValThrSerAlaThrGlyThrIle 839

2977 ......ACCTTG.........GCTGTCAAC.........AATACCGGCAA 3002
           |||:::         |||||||||         |||||||||::
 840 GlyGlyThrIleSerGlyAsnAlaValAsnValThrAlaAsnThrGlyAs 856

3003 CGAACCCGTA............................AGTCTCGAGC 3022
     :      |||                            :::
 856 pLeuThrValGluAspAlaAlaLysIleAspAlaThrGlyGlyAlaAlaT 873

3023 AATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTT 3072
       ||||||::::::::::|||||   :::|||      |||:::::::::
 873 hrLeuThrAlaThrSerGlyLysLeuThrThrLysAlaSerSerSerIle 889

3073 AATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCA 3122
     ::: |||                           |||::::::   ::::
 890 Thr.Ser.........................AlaAsnAsnGlnValA 897

3123 GCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAAC 3172
     ::|||||||||||   :::|||   :::      |||::::::   |||
 897 snLeuSerAlaLysAspGlySerIleGlyGlyAsnIleAsnAlaAlaAsn 913

3173 AAGAGCTTTCCGACAAACTCGGCA 3196
        ::::::   |||:::::::|||
 914 ValThrLeuAsnThrThrGlyAla 921
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41723

seq_documentation_block:

```
ID   AAR41723 standard; Protein; 1536 AA.
XX
AC   AAR41723;
XX
DT   26-APR-1994  (first entry)
XX
DE   High molecular weight protein 1 (HMW1).
XX
KW   HMW; high molecular weight protein; virus; vaccine; influenza;
KW   epitope; immunity; haemophilus influenzae.
XX
OS   Haemophilus influenzae.
XX
PN   WO9319090-A.
XX
PD   30-SEP-1993.
XX
```

```
PF   16-MAR-1993;   93WO-US02166.
XX
PR   16-MAR-1992;   92GB-0005704.
XX
PA   (BARE/) BARENKAMP S J.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1993-320683/40.
DR   N-PSDB; AAQ49506.
XX
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties
XX
PS   Claim 3; Figure 2; 100pp; English.
XX
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.
XX
SQ   Sequence   1536 AA;


alignment_scores:
            Quality:  265.00              Length:   1297
              Ratio:  0.438                Gaps:     73
 Percent Similarity:  46.646   Percent Identity: 21.126

alignment_block:
 US-09-303-518D-653 x AAR41723   ..

 Align seg 1/1  to: AAR41723  from: 1  to: 1536

    178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
        |||   |||   ::::::||||||   ||||||   :::|||
    199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal........ 211

    228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
              :::|||:::|||   ::::::   :::
    212 .......AsnLeuIleGlyGlyLysValLysAsnGlu............. 221

    278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
          :::|||:::|||   ||||||      ::::::   |||||||||;::
    222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

    322 CAA.................................TATATTGTGAG 335
        :::                                 |||   :::::
    238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

    336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
        ::::   :::|||   :::   |||      |||:::   |||:::|||:
    254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271

    386 GCAAT.......................CCCGATCAGCACCGCTTTTCT 411
        ::|||                       :::|||   :::::|||
    271 lyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287

    412 TACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCC 461
          :::   |||:::::::::      |||:::|||
    288 AlaAspSerValSerLysAsp......LysSerGly.............. 297

    462 TTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATG 511
                                    ::::::|||   :::|
    298 ................................AsnIleValLeuSerA 303

    512 CAGAACCTGTTGAG.................ATGACCAGTTATATGGAT 543
        ||:::      |||               :::::::::      :::
    303 laLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319

    544 GGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT......CG 587

        |||   :::   |||            |||:::|||      ::
    320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLy 336

    588 AATCGGAGCA.........................GGCAGACAATATT 610
        :   ||||||                        |||:::   |||
    336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353

    611 GGCGGTCTGATGAA...GACGAACCCAATAAC.................. 639
              :::||||||      |||   :::|||
    353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369

    640 ..............CGCGAAAGTTCATATCATATTGCAAGC........ 666
                       :::   |||:::   ::::::::::::
    370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLy 386

    667 ..........GCATATTCTTGG............CTCGTCGGTGGCAATA 694
                  |||      |||            |||:::   ||||||
    386 sGlyGlyArgAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnI 403

    695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAA 744
          ||||||   |||||||||
    403 leAsnAlaGln...GlySerGly........................... 409

    745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
                                    :::      |||||||||
    410 .....................AspIleAlaLysThrGlyGlyPheValGl 419

    795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGT 844
          ||||||      :::||||||   |||
    419 uThrSerGlyHisAspLeuPheIleLysAsp................... 429

    845 TAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAAT 894
          |||::::::::::::::::
    430 .....AsnAlaIleValAspAla........................... 435

    895 GGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGG 944
                          |||:::|||:::   |||
    436 ..................LysGluTrpLeuLeuAsp............Ph 442

    945 AGATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTT 994
         |||:::   |||:::      |||         :::::::::
    442 eAspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerG 459

    995 TTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACAC 1044
          :::|||:::   :::|||:::|||                  :::
    459 luAspAspGluTyrThrGlySerGly..................AsnSer 469

   1045 TATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTTAATGT 1094
          |||   |||   |||   |||   :::|||:::   |||   |||::
    470 AlaSerThrProLysArgAsnLysGluLysThrThr...LeuThrAsnTh 485

   1095 TTCTTTATCCGAG.............................ACAGCAA 1114
        ::::|||::::::                             ||||||:
    485 rThrLeuGluSerIleLeuLysLysGlyThrPheValAsnIleThrAlaA 502

   1115 GAGAACCTGTTTAT................................. 1128
        :::::   :::|||
    502 snGlnArgIleTyrValAsnSerSerIleAsnLeuSerAsnGlySerLeu 518

   1129 ..................CATGCTGCAGGTGGGGTCAACAGTTATCGACC 1160
                           ::::::::::|||||||||
    519 ThrLeuTrpSerGluGlyArgSerGlyGlyGlyVal.............. 530

   1161 CAGACTGAATAAT.........................GGAGAAAATA 1183
         ::::::||||||                        |||   |||:
    531 .GluIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsnL 547

   1184 TTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATC 1233
        ::::::::   :::   |||      :::      :::   :::||||||
    547 euThrIleTyrSerGlyGlyTrpValAsp......ValHisLysAsnIle 561

   1234 AACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCC 1283
        :::   ||||||              :::|||||||:::::::::::::
```


562 SerLeuGlyAla.....GlnGlyAsnIleAsnIleThrAl 573
1284 TAAA.....AACAACGAAACGTGGCAAGGCG 1309
573 aLysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyG 590
1310 CGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGC 1359
590 lnGlyThr...IleThrSerGlyAsnGlnLysGlyPheArgPheAsnAsn 605
1360 GTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCA 1409
606 ValSer.....LeuAsnGlyThrGlySerGly...LeuGlnPheTh 618
1410 AGCCAAAGGGGAAAAC.....CAAGGCTCGG 1435
618 rThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyThrL 635
1436 TCAGCGTGGGCGACGGTAAAGTCATCTTA.....GAT 1467
635 euAsnIle...SerGlyLysValAsnIleSerMetValLeuProLysAsn 650
1468 CAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTT..... 1503
651 GluSerGlyTyrAspLysPheLysGlyArgThrTyrTrpAsnLeuThrSe 667
1504 .....AGTGAAATCGGC.....TTGGTCAGCGGCAGGG 1531
667 rLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerArgG 684
1532 GGACG.....GTG 1539
684 lySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyIle 700
1540 CAACTGAATGCCGATAATCAGTTCAACCCCGACAAA..... 1575
701 SerPheAsnLysAspThrThrPheAsnValGluArgAsnAlaArgValAs 717
1576 .....CTCT 1579
717 nPheAspIleLysAlaProIleGlyIleAsnLysTyrSerSerLeuAsnT 734
1580 ATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCG 1629
734 yrAlaSerPheAsnGlyAsnIleSerValSerGlyGlyGlySerValAsp 750
1630 TTCCACCGCATT.....CAAAATACCGATGAAGGGGCGATGATTGT 1670
751 PheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyValValIl 767
1671 CAACCACAAT.....CAAGACAAAG 1690
767 eAsnSerLysTyrPheAsnValSerThrGlySerSerLeuArgPheLysT 784
1691 AATCCACCGTTACCATTACAGGC.....AATAAAGATATTACT... 1728
784 hrSerGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeu 800
1729 ...ACAACCGGCAATAAC..... 1743
801 AsnAlaThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAspGl 817
1744 .....AACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGT. 1782
817 yMetIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGlyG 834
1783 .....TGGTTTGGCGAGAAAGATGCA...ACCAAAACGAACGGGCGG 1821
834 lyAsnIleThrPheGlySerArgLysAlaValThrGluIleGluGlyAsn 850
1822 CTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTA..... 1863
851 ValThrIleAsn.....AsnAsnAlaAsnValThrLeuIleGlySe 864
1863 1863
864 rAspPheAspAsnHisGlnLysProLeuThrIleLysLysAspValIleI 881
1864 .....CTGCTTTCCGGCGGA.....ACAAATTTAAACGGC 1893
881 leAsnSerGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGly 897
1894 AATATCACG...CAAACAAACGGCAAACTGTTTTTCAGCGGCA..... 1933
898 AsnLeuThrValGluSerAsn.AlaAsn...PheLysAlaIleThrAsnP 913
1934 .....GACCGACACCGCACGCCTACAATC 1957
913 heThrPheAsnValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIle 929
1958 ATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATC 2007
930 SerIleAlaLysGlyGlyAlaArgPheLysAspIleAspAsnSerLysAs 946
2008 GTGTGGGACAACGATTGGATCGACCGC.ACATTTAAAGCGGAAAACTTCC 2056
946 nLeuSerIleThrThrAsnSerSerSerThrTyrArg..... 958
2057 ATATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAA 2106
959 .....ThrIleIleSerGlyAsnIleThrAsnLysAsn 969
2107 GGCGATTGGCATTTAAGCAATCACGCC.....CAAGCAGTTTT 2144
970 GlyAspLeuAsnIleThrAsnGluGlySerAspThrGluMetGlnIleGl 986
2145 CGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGA 2194
986 yGlyAspValSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysI 1003
2195 CGGGTCTGACAAGTTGTACCGAAAAAACCATTACC...GACGATAAAGTG 2241
1003 leAsnIleThr.....LysGlnIleThrIleLysAlaGlyVal 1015
2242 ATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGA 2291
1016 AspGlyGluAsnSerAspSerAspAlaThrAsnAsnAlaAsnLeuThrIl 1032
2292 TCACGCT.....CATTTAAATCTCACAGGACTT. 2319
1032 eLysThrLysGluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPheA 1049
2320 .....GCCACACTCAACGGC 2334
1049 snLysAlaGluIleThrAlaLysAspGlySerAspLeuThrIleGlyAsn 1065
2335 AATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCAC 2384
1066 ThrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGlnVa 1082
2385 CCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATC 2434
1082 lLysAspSerLysIleSerAlaAspGlyHisLysValThrLeuHisSerL 1099
2435 AAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAAT 2484
1099 ysValGluThrSerGlySerAsnAsnAsnThrGluAspSerSerAspAsn 1115
2485 CTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGC 2534
1116 AsnAlaGlyLeuThrIleAspAlaLysAsnValThrValAsnAsnAsnIl 1132
2535 TAAGGCAAAC.....GTAAGCCATTCCGCACTCAACGGCAATGTC.... 2574
1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyGluIleThrT 1149

```
2575 ..........TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGC 2613
1149 hrLysThrGlyThrThrIleAsnAlaThrThrGlyAsnValGluIleThr 1165

2614 CGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACG.........GCATT 2654
1166 AlaGlnThrGlySerIleLeuGlyGlyIleGluSerSerSerGlySerVa 1182

2655 ACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCA 2704
1182 lThrLeuThrAlaThrGluGlyAlaLeu............AlaValSerA 1195

2705 ATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACAC 2754
1195 snIleSerGlyAsnThrValThrValThrAlaAsnSer............ 1207

2755 GATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCG 2804
1208 .........GlyAlaLeuThrThrLeuAlaGlySer.............. 1216

2805 TTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCC 2854
1217 .ThrIleLysGlyThrGluSerValThr......ThrSerSerGlnSerG 1231

2855 GTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA... 2901
1231 ly............AspIleGlyGlyThrIleSerGly...GlyThrVal 1242

2902 ...TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAA 2948
1243 GluValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLy 1259

2949 GCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTC.......... 2988
1259 s...AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleG 1275

2989 ..............AACAATACCGGCAACGAACCCGTAAGTCTCGAGCAA 3024
1275 lyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1291

3025 TTGACGGTA.......................GTGGAAGGAAAAGACAA 3050
1292 LeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaTh 1308

3051 CACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCG 3100
1308 rLeuThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleT 1325

3101 ATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGC 3150
1325 hrSer.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAl 1341

3151 CTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACA 3187
1341 aGlySerIleAsnAlaAlaAsnValThrLeuAsnThr 1353
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU34143

seq_documentation_block:

ID AAU34143 standard; Protein; 2086 AA.
XX
AC AAU34143;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #419.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS52002.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 5639; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2086 AA;

alignment_scores:

Quality: 261.00 Length: 1529
Ratio: 0.345 Gaps: 80
Percent Similarity: 49.444 Percent Identity: 22.041

alignment_block:

US-09-303-518D-653 x AAU34143 ..

Align seg 1/1 to: AAU34143 from: 1 to: 2086

```
289 TCGCGTAACGGCGTGGCGGCATTGGCGGGCGAT.............CAATA 326
  5 ThrAsnAsnGlyIleThrValThrAlaGlyThrPheAsnProAlaAspTh 21

327 TATTGTGAGCGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTG 376
 21 rIleGlnValValAlaThrGlnGly.....................SerG 31

377 CGGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTG... 423
 31 lyGluThrIleSerAspGluGlnArgSerAspAspPheThrValValAla 47

424 .........AAAAGAAATAATTATAAAGCAGGGACTAACGGCCAT..... 459
```


48 ProGlnProAsnHisAlaAsnThrLysIleTrpGlnAsnGlyHisIleAs 64
460 .....CCTTATGGCGGCGATTATCATATG..... 483
64 pIleThrProAsnAsnProSerGlyHisLeuIleAsnProThrGlnAlaM 81
484 .....CCGCGTTTGCACAAATTTGTCACAGATGCAGAACCTGTTGAG 525
81 etAspIleAlaTyrThrGluLysMetGlyAsnGlyAlaGluHisSerLys 97
526 ATGACCAGTTATATGGATGGG.....TGGAAATACGCTGATTT 563
98 ThrIleAsnValValArgGlyGlnAsnAsnGlnTrpThrIleAla..... 112
564 AAATAAATACCCTGATCGTGTTCGAATCGGAGCA.....GGCAGACAAT 607
113 .AsnLys...ProAspTyrValThrLeuAspAlaHisThrGlyLysValT 128
608 ATTGGCGGTCTGATGAAGACGAACCCAATAAC..... 639
128 hrPheAsnAlaAsnThrValLysProAsnSerAlaIleThrIleThrPro 144
640 CGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGG 689
145 LysAlaGlyThrGlyHisSerAlaSerSer..... 154
690 CAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTA 739
155 .AsnProSerThrLeuThrAlaProAlaThrHisThrValAsnThr...T 170
740 GCGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCA 789
170 hrGluIleValLys.....AspTyrGlySerAsnValThrAlaAlaGlu 184
790 TTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAA 839
185 IleAsnAsnAla.....ValGlnValAlaAsnLysArgTh 196
840 GTGGTTAATT...AATGGGGTATTGCAAACAGGCAACCCCTATATAGGAA 886
196 rAlaThrIleLysAsnGlyThrAlaMetProThrAsnLeuAlaGlyGlyS 213
887 AAAGCAATGGCTTC.....CAGCTAGTTCGTAAAGATTGGTTCTATGAT 930
213 erThrThrThrIleProValThrValThrTyrAsnAspGlySerThrGlu 229
931 GAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAA.. 978
230 GluVal.....GlnGluSerIlePheThrLysAlaAspLysAr 242
979 .....AATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAG 1018
242 gGluLeuIleThrAlaLysAsnHisLeuAspAspProValSerThrAspG 259
1019 GAAAA.....ATCGATGCCAAACATAAACAC 1044
259 lyLysLysProGlyThrIleThrGlnTyrAsnAsnAlaIleHisAsnAla 275
1045 TATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTTAATGT 1094
276 GlnGlnGlnIleAsnThrAlaLysThrGluAlaGlnGlnValIleAsnAs 292
1095 TTCTTTATCC.....GAGACAGCAAGAGAACCTGTTTATC 1129
292 nGluArgAlaThrProGlnGlnValSerAspAlaLeuThrLysValArgA 309
1130 ATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAA 1179
309 laAlaGlnThrLysIleAsnGluAlaLysAlaLeuLeuGlnAsnLysGlu 325
1180 AATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAA 1229
326 AspAsnSerGlnLeuValThrSerLysAsnAsn.....LeuGlnSerSe 340
1230 CATCAACCAA.....GGCGCGGGCGGTTTGTATTTTGAGGGT.....A 1267
340 rValAsnGlnValProSerThrThrGlyMetThrGlnGlnSerIleAspA 357
1268 ATTTTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTT 1317
357 snTyrAsnAlaLysLysArgGluAlaGluThrGluIleThrAlaAlaGln 373
1318 CATATCAGTGATGGCAGTACCGTTACTTGGAAA..... 1350
374 ArgValIleAspAsnGlyAspAlaThrAlaGlnGlnIleSerAspGluAs 390
1351 .GTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGC 1399
390 nThrThrAlaGlnArgAsnAspLeuThrAsnGlnIleSerGlnAlaThrA 407
1400 TGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGAC 1449
407 snLeuAlaAlaValGluSerValLysGlnSerAlaAsnSerLeu...Asp 422
1450 GGTAAAGTC.....ATCTTAGATCAG..... 1470
423 GlyAlaMetGlyAsnLeuGlnThrAlaIleAsnAspLysSerGlyThrLe 439
1471 .....CAGGCGGACGATCAAGGCAAAAAACAAGCCT 1501
439 uAlaSerGlnAsnPheLeuAspAlaAspGluGln...LysArgAsnAlaT 455
1502 TTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCC 1551
455 yrAsnGln.....AlaValSerAsnAlaGluThrIle...LeuAsnLys 468
1552 GATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACG 1601
469 GlnThrGlyProAsnThrAlaLys.....ThrAlaValGl 480
1602 TTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCG 1651
480 uGlnAlaLeuAsnAsnValAsnSerAlaLysHisAlaLeuAsnGlyThrG 497
1652 ATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTT 1701
497 ln.....AsnLeuAsnAsnAlaLysGlnAlaAlaIle 507
1702 ACC...ATTACAGGCAATAAAGATATTACTACAACCGGCAATAACAACAA 1748
508 ThrAlaIleAsnGlyAlaSerAspLeu.....AsnGlnLysGl 520
1749 CTTGGATAGCAAAAAAGAAATTGCC.....TACAACG 1780
520 nLysAspAlaLeuLysAlaGlnAlaAsnGlyAlaGlnArgValSerAsnA 537
1781 GTTGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCTCAATCTG 1830
537 laGlnAspValGlnArgAsnAlaThrGluLeuAsnThrAlaMetGly... 552
1831 AATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAAC 1880
553 ThrLeuLysHisAlaIleAlaAspLysThrAsnThrLeuAlaSerSerLy 569
1881 AAATTTAAACGGCAATATCACGCAA.....A 1906
569 sTyrValAsnAlaAspSerThrLysGlnAsnAlaTyrThrThrLysValT 586
1907 CAAACGGCAAACTGTTTTTCAGCGGCAGACCGACA..... 1941
586 hrAsnAlaGluHisIleIleSerGlyThrProThrValValThrThrPro 602
1942 .....CCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAAT 1982
603 SerGluValThrAlaAlaAlaAsnGlnValAsnSerAlaLysGlnGluLe 619

```
 1198 nLysTyrArgAspAlaGlu...GlnSerLysLysThrAlaTyrAspGlnA 1214

 3485 CCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCC 3534
      || :::|||   |||           |||::::::   :::   |||
 1214 la.ValAlaAlaAlaLysAlaIleLeuAsnLysGlnThrGlySerAsnSe 1230

 3535 CAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGA 3584
       :::::::::|||:::      |||:::::::   :::
 1230 rAspLysAlaAlaValAspArgAlaLeuGlnGlnVal............. 1242

 3585 ATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACC 3634
                                          |||:::
 1243 .....................................ThrSerThrLys 1246

 3635 GCGTGTTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGG 3684
        :::|||       |||           ::::::|||||||:::
 1247 AspAlaLeuAsnGlyAspAlaLysLeuAlaGluAlaLysAlaAlaAla.. 1262

 3685 GACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAAC 3734
                           |||   :::      |||   |||
 1263 ......................LysGlnAsnLeuGlyThrLeuAsnHisI 1272

 3735 CGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCG 3784
        |||   |||            ::::::|||:::      |||
 1272 leThrAsnAla............GlnArgThrAlaLeuGluGly..... 1282

 3785 GCATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACGGCATC 3834
                                        :::   :::   ::
 1283 ..................................GlnIleAsnGlnAl 1287

 3835 GGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCAT 3884
      ::::|||      |||:::   ||||||      ::::::||||||   :
 1287 aThrThrValAspGlyValAsnThrValLysThrAsnAlaAsnThrLeuA 1304

 3885 CGGCAGGTTCGACATCGGCATCAGCG............CGGGCGCGGGTT 3922
      ::::::::   :::|||   ::::::            :::      :::
 1304 spGlyAlaMetAsnSerLeuGlnGlySerIleAsnAspLysAspAlaThr 1320

 3923 TTAGTAGCGGCAGCCTTTCAGACGGCATCAGAGGCAAAA...TCCGCCGC 3969
      |||            :::      |||:::|||:::|||   :::|||
 1321 LeuArgAsnGlnAsnTyrLeuAspAlaAspGluSerLysArgAsnAlaTy 1337

 3970 CGCGTGCTGCATTACGGCATTCAGG.....................C 3995
       :::   ::::::||||||   :::
 1337 rThrGlnAlaValThrAlaAlaGluGlyIleLeuAsnLysGlnThrGlyG 1354

 3996 AAGATACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGC...ACATCG 4042
        :::|||:::::::::::::::|||:::|||   |||      |||
 1354 lyAsnThrSerLysAlaAspValAspAsnAlaLeuAsnThrValThrArg 1370

 4043 GCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTC 4092
      |||:::   |||:::::::::   :::::::::   :::||||||||||
 1371 AlaLysAlaAlaLeuAsnGlyAlaGluAsnLeuArgAsnThrLysThrSe 1387

 4093 AATATCGCCACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAA 4142
      |                         :::|||   |||   :::::::|||
 1387 r.....................AlaThrAsnThrIleAsnGlyLeuP 1396

 4143 GGCAGATTATTCATTCAAACCGGCGCAAC 4171
        ::::::   :::   ::::::   |||
 1396 roAsnLeuThrGlnLeuGlnLysAspAsn 1405

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37018

seq_documentation_block:
ID   AAU37018 standard; Protein; 3158 AA.
XX
AC   AAU37018;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #1188.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS54877.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 12611; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   3158 AA;


alignment_scores:
           Quality:  261.00              Length:   1107
             Ratio:  0.466                 Gaps:     51
 Percent Similarity:  50.587   Percent Identity: 21.951

alignment_block:
 US-09-303-518D-653 x AAU37018   ..

 Align seg 1/1  to: AAU37018  from: 1  to: 3158

    1144 GTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTAT 1193
         |||||||:::   :::       ||||||:::::::::|||::::::
     818 ValAsnAlaAlaLysGlnAlaLeuAsnGlyAsnAspAsnLeuAlaAsnAl 834

    1194 T...GACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATCAACCAAG 1240
              ::::::   |||   :::|||         :::::::::|||:::
     834 aLysGlnGlnProLysGlnGlnLeuAlaAsnLeuThrHisLeuAsnAsp. 850
```

```
1241 GCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCCTAAAAAC 1290
      |||           ||||||:::::::::|||   :::|||
 851 ..AlaGlnLysGlnSerPheGluSerGlnIleThrGlnAlaProLeu... 865

1291 AACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGGCAGTACCGT 1340
                                 :::::::|||   :::|||::
 866 ............................ValThrAspValThrThrIl 872

1341 TACTTGGAAAGTA.............................AACG 1357
     ::::   |||:::                            |||:
 872 eAsnGlnLysAlaGlnAlaLeuAspHisAlaMetGluLeuLeuArgAsnS 889

1358 GCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTT 1407
     :::::|||:::::::::   :::      ::::::            :::
 889 erIleAlaAspAsnGlnAlaThrLeuAlaSerGluAspTyrHisAspAla 905

1408 CAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGT 1457
        |||:::   :::|||:::   :::   :::|||      :::|||
 906 ThrAlaGlnArgGlnAsnAspTyrAsnGlnAlaValThrAlaAlaLysAs 922

1458 CATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTG 1507
      |||::::::|||   ::::::         :::   :::      ::::
 922 nIleIleAsnGlnThrThrSerProThrMetAsnProAspGluValAsnA 939

1508 AAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAAT 1557
     ::               ::::::      |||   |||::::::|||:::
 939 rgAlaThrThrGlnValAsnAsnThrLysValAlaLeuAspGlyAspGlu 955

1558 CAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGA 1607
     ::::::         |||         :::   |||   ::::::|||||
 956 AsnLeuAlaAlaAlaLysGlnGlnAlaAsnAsnArgLeuAsnGlnLeuAs 972

1608 T...TTGAACGGG.........CATTCGCTTTCGTTCCACCGCATTCAAA 1645
     |   ||||||:::         ::::::|||:::   :::      |||:
 972 pHisLeuAsnAsnAlaGlnLysGlnGlnLeuGlnSerGlnIleAlaGlnS 989

1646 ATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCC 1695
     :::::|||   :::|||   :::   :::|||:::|||      ||||||
 989 erSerAspIleAlaAla...ValAsnGlyHisLysGlnThrAlaGluSer 1004

1696 ACCGTTACCATTACAGGCAAT.....................AAAGA 1721
           |||      ||||||                     :::
1005 LeuAsnThrAlaMetGlyAsnLeuIleAsnAlaIleAlaAspHisGlnAl 1021

1722 TATTACTACAACCGGCAATAACAACAACTTGGATAGCAAAAAAGAAATTG 1771
      :::         ||||||     |||   |||:::   |||:::   |
1021 aValGluGlnArgGlyAsnPheIleAsnAlaAspThrAspLysGlnThrA 1038

1772 CCTACAACGGTTGGTTTGGCGAGAAAGATGCA......ACCAAAACGAAC 1815
     |||||:::         :::|||   :::|||      :::|||   :::
1038 laTyrThrThrAlaValAsnGluAlaGluAlaMetIleAsnLysGlnThr 1054

1816 GGGCGGCTCAATCTGAATTACCAACCGGAA..................GA 1847
     |||:::   |||   |||   |||||||||                  ::
1055 GlyGln...AsnAlaAsn...GlnProGluValGluGlnAlaIleThrLy 1069

1848 AGCGGATCGCACTTTA...CTGCTTTCCGGCGGAACAAATTTA....... 1887
     :::::::   ||||||      |||:::|||      ||||||
1069 sValGlnThrThrLeuGlnAlaLeuAsnGlyAspHisAsnLeuGlnValA 1086

1888 ..AACGGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGA 1935
       :::   |||   ||||||:::         |||         :::
1086 laLysThrAsnAlaThrGlnAlaIleAspAlaLeuThrSerLeuAsnAsp 1102

1936 CCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGA 1985
     |||         |||   :::         ::::::   :::   :::
1103 ProGlnLysThrAlaLeuLysAspGlnValThrAlaAlaThrLeuValTh 1119

1986 AGGTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCA 2035
      ::::::   |||                          |||:::::::
1119 rAlaValHisGln.........................IleGluGlnA 1127

2036 CATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTTCC 2085
```

```
     ::   ::::::   |||      :::::::|||   :::      ::::::
1127 snAlaAsnThrLeuAsnGlnAlaMetHisGlyLeuArgGluSerIleGln 1143

2086 CGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGC........... 2124
        |||:::|||         ::::::   :::::::::
1144 AspAsnAlaAlaThrLysAlaAsnSerLysTyrIleAsnGluAspGlnPr 1160

2125 ..........AATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAA 2164
             |||:::   ||||||||                 :::   :
1160 oGluGlnGlnAsnTyrAspGlnAlaVal..............GlnAlaA 1172

2165 GCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACC 2214
     :::::::::|||   :::::::::      |||
1172 laAsnSerIleIleAsnGluGlnThrAlaThr............LeuAsp 1184

2215 GAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGA 2264
     :::::::::|||::::::      :::   ::::::   :::|||:::
1185 AsnAsnAlaIleAsnGlnAlaAlaThrThrValAsnThrThrLysAlaAl 1201

2265 CATCAGAGGCAATGTCAGCCTT.........GCCGATCACGCTCATTTAA 2305
      ::::::|||:::|||:::|||            |||||||||      :
1201 aLeuHisGlyAspValLysLeuGlnAsnAspLysAspHisAlaLysGlnT 1218

2306 ATCTCACAGGACTTGCCACACTCAACGGC..................... 2334
     ::::::::   ||||||   ||||||:::
1218 hrValSerGlnLeuAlaTyrLeuAsnAsnAlaGlnLysHisMetGluAsp 1234

2335 AATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCAC 2384
     :::|||      :::      |||:::   :::|||   ::::::   ||
1235 ThrLeuIleAspSerGluThrThrArgIleAlaValLysGlnAspLeuTh 1251

2385 CCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATC 2434
     |:::         |||:::   :::   |||         |||:::::::|
1251 rGluAlaGlnAlaLeuAspGlnLeuMetAsn.........ThrLeuGlnG 1265

2435 AAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAAT 2484
     ||:::      :::   :::::::::   :::::::::|||      |||
1265 lnSerIleAlaAspLysAspAlaThrArgAlaSerSerAlaTyrValAsn 1281

2485 CTAAGCAACAAC.....................GCCGTACAAAACGGCAG 2513
        :::   |||                     ||||||||||||
1282 AlaGluProAsnLysLysGlnAlaTyrAspGluAlaValGlnAsn..... 1296

2514 TCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCA 2563
               :::::::::      |||:::::::::::::   :::::::|
1297 ..........AlaGluSerIleIleAlaGlyLeuAsnAsnProThrIleA 1310

2564 AC...GGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAAC 2610
     ||   ||||||||||||   |||   :::||||||          :::|||
1310 snLysGlyAsnValSerSerAlaThrGlnAlaValThrThrSerLysAsn 1326

2611 AGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACACTT 2660
                                    |||   |||   :::   :::||
1327 ............................GlyLeuAspGlyValGluArgLe 1334

2661 AAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATTTAA 2710
     |   :::::::::   |||      :::|||::::::|||::::::|||:
1334 uAlaGlnAspLysGlnThr......AlaGlyAsnSerLeuAsnHisLeuA 1349

2711 AC............................CTTGACAACGCCACCATTACA 2733
     ::                           |||:::|||
1349 spGlnLeuThrProAlaGlnGlnGlnAlaLeuGluAsn.........Gln 1362

2734 CTCAATTCCGCCTATCGACACGATGCGGCAGGC.........GCGCAAAC 2774
     :::|||:::|||      :::|||   ::::::         ||||||::
1363 IleAsnAsnAlaThrThrArgAspLysValAlaGluIleIleAlaGlnAl 1379

2775 CGGCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTAT 2824
     :   :::      :::|||         ::::::   ::::::   ::::
1379 aGlnAlaLeuAsnGluAla......MetLysAlaLeuLysGluSerIleL 1394

2825 CCGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTA 2874
     ::      |||   |||:::|||:::|||:::|||   :::
```

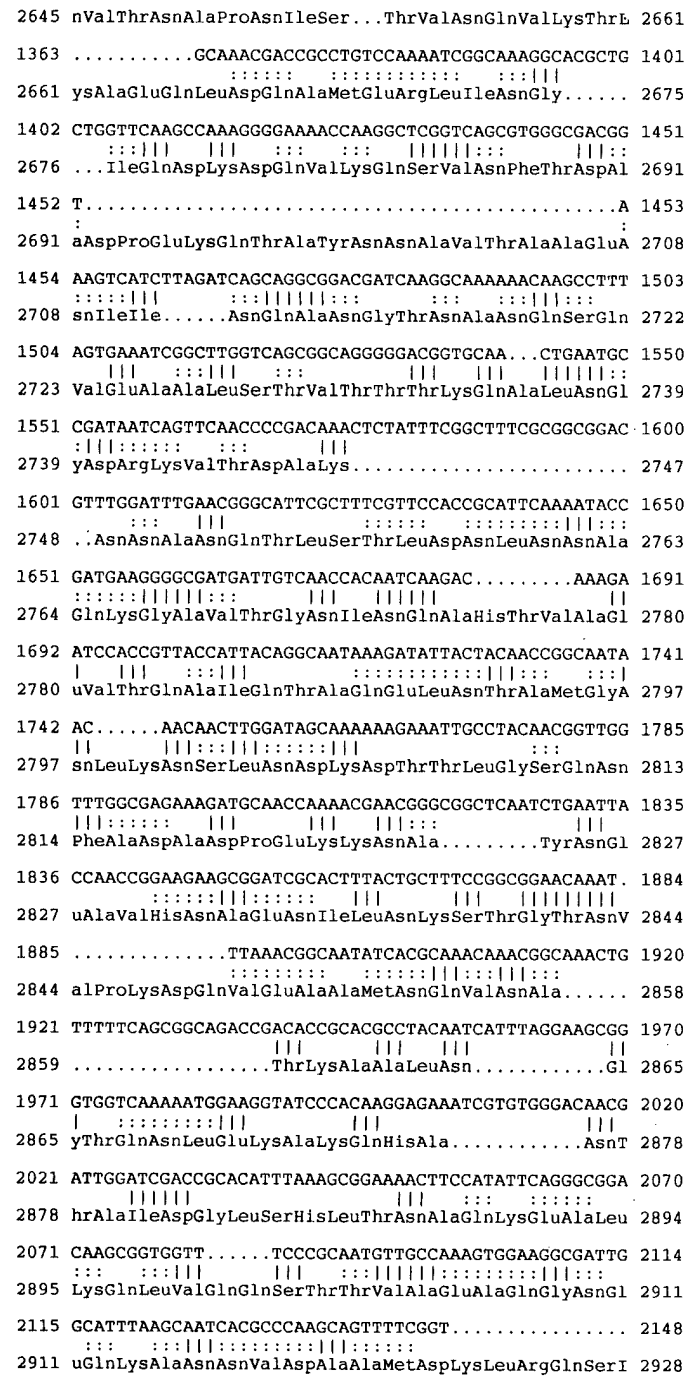
```
1394 ysAspGlnProGlnThrGluAlaSerSerLysPheIleAsnGluAspGln 1410

2875 AACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTT 2924
          |||              :::   :::      :::::::|||::
1411 AlaGlnLysAspAlaTyrThrGlnAlaValGlnHisAlaLysAspLeuIl 1427

2925 CGGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTT 2974
     ::::   ::::::           |||:::::|||
1427 eAsn...LysThrThrAspProThrLeuValLysSerValIleAspGlnA 1443

2975 ACACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCTCGAGCAA 3024
       |||   |||||||||:::::   |||:::      :::   ::::::
1443 laThrGlnAlaValAsnAspAlaLysAsnAsnLeuHisGlyAspGlnLys 1459

3025 TTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAA 3074
     |||           ::::::|||:::          :::|||:::|||||
1460 Leu............AlaGlnAspLysGlnArgAlaThrGluThrLeuAs 1472

3075 TTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCAGC 3124
     |   :::|||:::|||   ::::::
1472 n...AsnLeuSerAsnLeuAsnThrProGln..................... 1481

3125 TTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAA 3174
          |||:::      |||   ::::::::::|||   :::   :::
1482 .....ArgGlnAlaLeuGluAsnGlnIleAsnAsnAlaAlaThrArgGly 1496

3175 GAGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGAC 3224
     |||::::::::::||||||   :::|||          :::   |||:::
1497 GluValAlaGlnLysLeuThrGluAlaGlnAlaLeuAsnGlnAlaMetGl 1513

3225 GGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCGGAA.......... 3264
      |||   :::   ::::::      :::|||||||:::|||
1513 uAlaLeuArgAsnSerIleGlnAspGlnGlnGlnThrGluAlaGlySerL 1530

3265 ...........AAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCC 3303
                :::|||:::   |||:::   ||||||      ||||||
1530 ysPheIleAsnGluAspLysProGlnLys...AspAlaTyrGlnAlaAla 1545

3304 GGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCA 3353
        ::::::|||   :::      ::::::::::::::|||:::   ::
1546 ValGlnHisAlaLysAspLeuIleAsnGlnThrSerAsnProThrLeuAs 1562

3354 GGCAGGCGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAAC 3403
     :   :::                        |||:::|||:::      :::
1562 pLysAla.....................GlnValGluGlnLeuThrGlnG 1572

3404 GGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAA 3453
       |||:::      ||||||:::      :::   |||:::   |||:::
1572 lyValAsnGlnAlaLysAspAsnLeuHisGlyAspGlnLysLeuAlaAsp 1588

3454 ACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGA 3503
        :::                          ::::::|||      ||
1589 AspLys.........................GlnHisAlaValThrAs 1596

3504 TTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGA 3553
     ||||   |||   :::         :::||||||   |||      |||
1596 pLeuAsnGlnLeuAsnSerLeuAsnAsnProGlnArgGlnAla...LeuG 1612

3554 TCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAAC 3603
       |||:::      |||:::
1612 luSerGlnIleAsnAsnAla............................. 1618

3604 AGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAA..... 3648
               |||::::::||||||:::   :::   :::||||||
1619 .........AlaThrArgAspGluValAlaGlnLysLeuAlaGluAlaGl 1632

3649 .......GACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCA 3691
            |||:::        :::      :::::::|||:::|||   :
1632 nAlaLeuAspGlnAlaMetGlnAlaLeuArgAsnSerIleGlnAspGlnG 1649

3692 AACACTACCGT......TCGCAAGATTTCCGCGCCTACCGCCAACAAACC 3735
     :::::   :::      |||:::   ::::::      :::   |||
1649 lnGlnThrGluSerGlySerLysPheIleAsnGluAspLysProGlnLys 1665

3736 GACCTGCGCCAAATCGGTATGCAG......AAAAACCTCGGCAGCGGGCG 3779
          |||      |||   :::::::|||      |||:::|||
1666 AspAlaTyrGlnAlaAlaValGlnHisAlaLysAspLeu........... 1678

3780 CGTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCG...... 3823
                                  |||:::|||||||| |||:::
1679 ...................IleAsnGlnThrGlyAsn.ProThrLeuAsp 1688

3824 ................................ACGACGGCATCGGCAAC 3840
                                     |||||||||:::   ::
1689 LysSerGlnValGluGlnLeuThrGlnAlaValThrThrAlaLysAspAs 1705

3841 TCGGCACGGC..................................... 3850
     :   ||||||
1705 nLeuHisGlyAspGlnLysLeuAlaArgAspGlnGlnGlnAlaValThrT 1722

3851 ..........TTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGCA 3889
                |||||:::::::      :::   :::   |||   :::
1722 hrValAsnAlaLeuProAsnLeuAsnHisAlaGlnGlnGlnAlaLeuThr 1738

3890 GGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTT 3939
        :::   :::|||   |||   :::|||:::   ||||||
1739 AspAlaIleAsnAla...AlaProThrArgThrGluValAlaGlnHisVa 1754

3940 TCAGACGGCATCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCAT 3989
      |||||||||:::|||         |||             :::   ::::
1754 lGlnThrAlaThrGluLeuAspHisAla............MetGluThrL 1767

3990 TCAGGCAAGATACCGCGCAGGTTTCGGCGGATTCGGCATCG...AACCGC 4036
     ::::::::   :::   ||||||:::::|||:::|||:::   |||
1767 euLysAsnLysValAspGlnValAsnThrAspLysAlaGlnProAsnTyr 1783

4037 ACATCGGCG.....................CAACGCGCTATTTCGTC 4062
     |||:::|||                    :::::::|||:::::::
1784 ThrGluAlaSerThrAspLysLysGluAlaValAspGlnAlaLeuGlnAl 1800

4063 CAAAAAGCGGATTACCGAT 4081
     :   :::   |||||||||
1800 aAlaGluSerIleThrAsp 1806
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37403

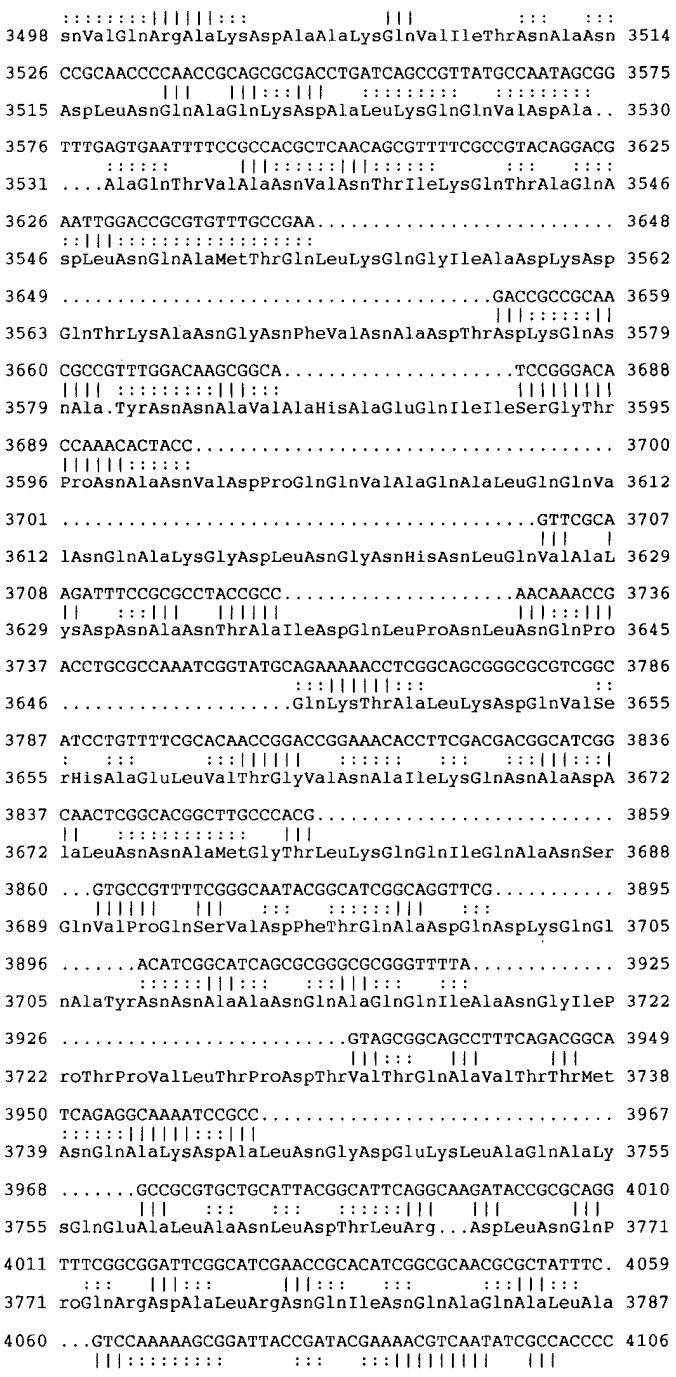
seq_documentation_block:

ID AAU37403 standard; Protein; 6281 AA.
XX
AC AAU37403;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #1573.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX

```
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS55262.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 12996; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   6281 AA;


alignment_scores:
            Quality:  260.50              Length:   1733
              Ratio:  0.333                Gaps:     80
 Percent Similarity:  45.124   Percent Identity: 19.792

alignment_block:
 US-09-303-518D-653 x AAU37403   ..

 Align seg 1/1  to: AAU37403  from: 1  to: 6281

     163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATAT 212
         :::|||   :::   ||||||   |||   ||||||
    2334 LysAspHisGlyAlaAsnLysAspLysGlnAlaValGluGlnAlaIleGl 2350

     213 TGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGC.............A 247
         :::|||   :::   ::::::   |||   |||             :
    2350 nSerValThrSerThrGluAsnAlaLeuAsnGlyAspAlaAsnLeuGlnA 2367

     248 AATCGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAAC 297
         :::::   |||:::|||   :::         :::::::::      |||
    2367 rgAlaLysThrGluAlaIleGlnAlaIleAspAsnLeuThrHisLeuAsn 2383

     298 GGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAA 347
                     |||      |||   :::   ::::::   :::
    2384 ThrProGlnLysThrAlaLeuLysGlnGlnValAsnAlaAlaGlnArgVa 2400

     348 CGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCGATC 397
         :::|||   :::::::::      :::|||   :::   |||       :
    2400 lSerGlyValThrAspLeuLysAsnSerAlaThrSerLeuAsnAsnAlaM 2417

     398 AGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAAGCAGGG 447
         ::   :::::::::      |||:::   :::::::::      |||
    2417 etAspGlnLeuLysGlnAlaIleAlaAspHisAspThrIleValAla... 2432

     448 ACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCACAA 497
                         :::|||:::|||
    2433 ..................SerGlyAsnTyr..................... 2436

     498 ATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATGGGT 547
                                                  |||:::|||:::|||   ::::::   :::|||   |||::::
    2437 .......ThrAsnAlaSerProAspLysGlnGlyAlaTyrThrAspAlaT 2451

     548 GGAAATACGCTGAT......TTAAATAAATACCCTGATCGTGTTCGAATC 591
         :::::   |||         ::::||      |||:::   :::
    2451 yrAsnAlaAlaLysAsnIleValAsnGlySerProAsnValIleThrAsn 2467

     592 GGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAACCCAATAACCG 641
         :::|||                      |||            ::::::||
    2468 AlaAla.....................AspValThrAlaAlaThrGlnAr 2477

     642 CGAAAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCA 691
         |   ::::::
    2477 gValAsnAsn..................................... 2480

     692 ATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGC 741
                |||::::::|||   :::|||   |||   ||||||::::::
    2481 ........AlaGluThrGlyLeuAsnGlyAspThr...AsnLeuAlaThr 2493

     742 GAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATT 791
            |||   :::   :::
    2494 AlaLysGlnGlnAlaLys............................... 2499

     792 TGGCGACAGTGGCTCACCAATG...TTTATCTATGATGCCCAAAAGCAAA 838
             |||:::          |||      :::   |||||||||||||||:
    2500 ....AspAlaLeuArgGlnMetThrHisLeuSerAspAlaGlnLysGlnS 2515

     839 AGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGA... 885
         ::      |||:::|||   ::::::::::::::          |||
    2515 er......IleThrGlyGlnIleAspSerAlaThrGlnValThrGlyVal 2529

     886 .........AAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTA 926
                  |||:::|||:::   :::|||
    2530 GlnSerValLysAspAsnAlaThrAsnLeu................... 2539

     927 TGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATC 976

    2539 ..................................................... 2539

     977 AAAATGGGAAATACTTTTTTAACGACAAT.................AAT 1008
                                  ||||||                |||
    2540 ......................AspAsnAlaMetAsnGlnLeuArgAsn 2548

    1009 AATGGCGCAGGAAAAATCGATGCCAAACATAAACACTATTCTCTACCTTA 1058
         :::   |||:::|||   |||:::|||            |||   |||||
    2549 SerIleAlaAsnLysAspAspValLys.........AlaSerGlnProTy 2562

    1059 TAGATTAAAAACACGAACCGTTCAA...TTGTTTAATGTTTCTTTA.... 1101
         |            |||      |||      :::|||::::::::
    2562 rValAspAlaAspArgAspLysGlnAsnAlaTyrAsnThrAlaValThrA 2579

    1102 ........................TCCGAGACAGCAAGAGAACCT... 1122
                                    |||:::   :::   :::|||
    2579 snAlaGluAsnIleIleAsnAlaThrSerGlnProThrLeuAspProSer 2595

    1123 ...GTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAA 1169
            |||   :::|||||||:::   |||::::::   :::      |||||
    2596 AlaValThrGlnAlaAlaAsnGlnValSerThrAsnLysThrAlaLeuAs 2612

    1170 TAATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATAC 1219
         |:::::::::|||:::::          :::   |||   |||
    2612 nGlyAlaGlnAsnLeuAla.........AsnLysLysGlnGlu......T 2624

    1220 TTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAAT 1269
           |||:::|||||||||||||
    2624 hrThrAlaAsnIleAsnGln............................. 2630

    1270 TTTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCA 1319
               :::|||      ||||||          |||      ::::::::
    2631 ......LeuSerHisLeuAsnAsnAlaGlnLysGlnAspLeuAsnThrGl 2645

    1320 TATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTG....... 1362
         :::::::::::::   :::::::::      ||||||   |||
```


2645 nValThrAsnAlaProAsnIleSer...ThrValAsnGlnValLysThrL 2661
1363 ..........GCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTG 1401
2661 ysAlaGluGlnLeuAspGlnAlaMetGluArgLeuIleAsnGly..... 2675
1402 CTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGG 1451
2676 ...IleGlnAspLysAspGlnValLysGlnSerValAsnPheThrAspAl 2691
1452 T.....A 1453
2691 aAspProGluLysGlnThrAlaTyrAsnAsnAlaValThrAlaAlaGluA 2708
1454 AAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTT 1503
2708 snIleIle......AsnGlnAlaAsnGlyThrAsnAlaAsnGlnSerGln 2722
1504 AGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAA...CTGAATGC 1550
2723 ValGluAlaAlaLeuSerThrValThrThrThrLysGlnAlaLeuAsnGl 2739
1551 CGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGAC 1600
2739 yAspArgLysValThrAspAlaLys..... 2747
1601 GTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACC 1650
2748 ..AsnAsnAlaAsnGlnThrLeuSerThrLeuAspAsnLeuAsnAsnAla 2763
1651 GATGAAGGGGCGATGATTGTCAACCACAATCAAGAC.....AAAGA 1691
2764 GlnLysGlyAlaValThrGlyAsnIleAsnGlnAlaHisThrValAlaGl 2780
1692 ATCCACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGCAATA 1741
2780 uValThrGlnAlaIleGlnThrAlaGlnGluLeuAsnThrAlaMetGlyA 2797
1742 AC......AACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGG 1785
2797 snLeuLysAsnSerLeuAsnAspLysAspThrThrLeuGlySerGlnAsn 2813
1786 TTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAATTA 1835
2814 PheAlaAspAlaAspProGluLysLysAsnAla.....TyrAsnGl 2827
1836 CCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAAT. 1884
2827 uAlaValHisAsnAlaGluAsnIleLeuAsnLysSerThrGlyThrAsnV 2844
1885 .....TTAAACGGCAATATCACGCAAACAAACGGCAAACTG 1920
2844 alProLysAspGlnValGluAlaAlaMetAsnGlnValAsnAla..... 2858
1921 TTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGG 1970
2859 .....ThrLysAlaAlaLeuAsn.....Gl 2865
1971 GTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACG 2020
2865 yThrGlnAsnLeuGluLysAlaLysGlnHisAla.....AsnT 2878
2021 ATTGGATCGACCGCACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGA 2070
2878 hrAlaIleAspGlyLeuSerHisLeuThrAsnAlaGlnLysGluAlaLeu 2894
2071 CAAGCGGTGGTT......TCCCGCAATGTTGCCAAAGTGGAAGGCGATTG 2114
2895 LysGlnLeuValGlnGlnSerThrThrValAlaGluAlaGlnGlyAsnGl 2911
2115 GCATTTAAGCAATCACGCCCAAGCAGTTTTCGGT..... 2148
2911 uGlnLysAlaAsnAsnValAspAlaAlaMetAspLysLeuArgGlnSerI 2928
2149 .....GTCGCACCGCATCAAAGCCACACAATCTGTACA 2181
2928 leAlaAspAsnAlaThrThrLysGlnAsnGlnAsnTyrThrAspAlaSer 2944
2182 CGT.....TCGGACTGGACGGGTCT 2201
2945 GlnAsnLysLysAspAlaTyrAsnAsnAlaValThrThrAlaGlnGlyIl 2961
2202 GACAAGTTGTACCGAAAAAACCATTACCGACGATAAAGTGATT..... 2244
2961 eIleAspGlnThrThrSerProThrLeuAspProThrValIleAsnGlnA 2978
2245 ..........GCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGC 2283
2978 laAlaGlyGlnValSerThrThrLysAsnAlaLeuAsnGlyAsnGluAsn 2994
2284 CTT.....GCCGATCACGCTCATTTAAATCTCACAGGACTTGCCAC 2324
2995 LeuGluAlaAlaLysGlnGlnAlaSerGlnSerLeuGlySerLeuAspAs 3011
2325 ACTCAACGGC.....AATCTTAGTG 2344
3011 nLeuAsnAsnAlaGlnLysGlnThrValThrAspGlnIleAsnGlyAlaH 3028
2345 CAGGCGGAGACACGCACTATACGGTTACGCGCAACGCC..... 2382
3028 isThrValAspGluAlaAsnGlnIleLysGlnAsnAlaGlnAsnLeuAsn 3044
2383 ACCCAAAACGGCAACCTC...AGCCTCGTGGGCAATGCCCAAGCAACATT 2429
3045 ThrAlaMetGlyAsnLeuLysGlnAlaIleAlaAspLysAspAlaThr.. 3060
2430 TAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCAT 2479
3061 ....LysAlaThrVal......AsnPheThrAspAlaAspGlnAlaLysG 3074
2480 TTAATCTAAGCAACAACGCCGTACAAAAC..... 2508
3074 lnGlnAlaTyrAsnThrAlaValThrAsnAlaGluAsnIleIleSerLys 3090
2509 .....GGCAGTCTGACGCTTTCCGACAACGCTAAGGCA..... 2541
3091 AlaAsnGlyGlyAsnAlaThrGlnAlaGluValGluGlnAlaIleLysGl 3107
2542 ....AACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTA..... 2580
3107 nValAsnAlaAlaLysGlnAlaLeuAsnGlyAsnAlaAsnValGlnHisA 3124
2580 2580
3124 laLysAspGluAlaThrAlaLeuIleAsnSerSerAsnAspLeuAsnGln 3140
2581 ......GCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGG 2624
3141 AlaGlnLysAspAlaLeuLysGlnGlnValGlnAsnAlaThrThrValAl 3157
2625 AAAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAA..... 2664
3157 aGlyValAsnAsnValLysGlnThrAlaGlnGluLeuAsnAsnAlaMetT 3174
2665 .....GACAGCGAATGGACGCTGCCGTCGGGC 2691
3174 hrGlnLeuLysGlnGlyIleAlaAspLysGluGlnThrLysAlaAspGly 3190
2692 ACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTC 2741
3191 .....AsnPheValAsnAlaAspProAspLysGlnAs 3201
2742 CGCCTATCGACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATG 2791
3201 nAlaTyrAsnGlnAlaValAlaLysAlaGlu..... 3211


2792 CGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCG..... 2835
3212 .....AlaLeuIleSerAlaThrProAspVal 3220
2836 .....CCAACTTCGGCAGAATCCCGTTTCAACACGCTGACG...GT 2873
3221 ValValThrProSerGluIleThrAlaAlaLeuAsnLysValThrGlnAl 3237
2874 AAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCT 2923
3237 aLysAsnAspLeuAsnGlyAsnThrAsnLeuAlaThrAlaLysGlnAsnV 3254
2924 TCGGCTACCGCAGCGGCAAATTG...AAGCTGGCGGAAAGTTCCGAAGGC 2970
3254 alGlnHisAlaIleAspGlnLeuProAsnLeuAsnGlnAlaGlnArgAsp 3270
2971 ACTTACACCTTGGCTGTCAACAATACCGGCAACGAACCC.....GTAAG 3014
3271 GluTyrSerLysGlnIleThrGlnAlaThrLeuValProAsnValAsnAl 3287
3015 TCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCG 3064
3287 aIleGlnGlnAlaAlaThrThrLeuAsnAspAlaMetThrGlnLeuLysG 3304
3065 AAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGG 3114
3304 lnGly.....IleAlaAsnLysAlaGlnIleLysGlySerGlu 3316
3115 CGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGT 3164
3317 AsnTyrHisAspAlaAspThrAspLysGlnThrAlaTyrAspAsnAlaVa 3333
3165 CAAAGAACAAGAG..... 3177
3333 lThrLysAlaGluGluLeuLeuLysGlnThrThrAsnProThrMetAspP 3350
3178 .....CTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCC 3219
3350 roAsnThrIleGlnGlnAlaLeuThrLysValAsnAspThrAsnGlnAla 3366
3220 TTGACGGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCGGAA..... 3264
3367 LeuAsnGlyAsnGlnLysLeuAlaAspAlaLysGlnAspAlaLysThrTh 3383
3264 3264
3383 rLeuGlyThrLeuAspHisLeuAsnAspAlaGlnLysGlnAlaLeuThrT 3400
3265 .....AAAGAC 3270
3400 hrGlnValGluGlnAlaProAspIleAlaThrValAsnAsnValLysGln 3416
3271 AACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCAATGCC..... 3315
3417 AsnAlaGlnAsnLeuAsnAsnAlaMetThrAsnLeuAsnAsnAlaLeuGl 3433
3316 .....ACCGAAAAGGCAGAAAGTGTT.....GCCGAACCGG 3346
3433 nAspLysThrGluThrLeuAsnSerIleAsnPheThrAspAlaAspGlnA 3450
3347 CCCGGCAGGCAGGCGGGGAAAATGCC.....GGCATTATG 3381
3450 laLysLysAspAlaTyrThrAsnAlaValSerHisAlaGluGlyIleLeu 3466
3382 CAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTT 3431
3467 SerLysAlaAsnGlySerAsnAlaSerGlnThrGluValGluGlnAlaMe 3483
3432 GGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCC 3481
3483 t.....GlnArgValAsnGluAlaLysGlnAlaLeuAsnGlyAsnAspA 3498
3482 GCGCCCGCCGCGCCCGCCGGGATTTGCCGCAA.....CCGCAGCCCCAA 3525


3498 snValGlnArgAlaLysAspAlaAlaLysGlnValIleThrAsnAlaAsn 3514
3526 CCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGG 3575
3515 AspLeuAsnGlnAlaGlnLysAspAlaLeuLysGlnGlnValAspAla.. 3530
3576 TTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACG 3625
3531 ....AlaGlnThrValAlaAsnValAsnThrIleLysGlnThrAlaGlnA 3546
3626 AATTGGACCGCGTGTTTGCCGAA..... 3648
3546 spLeuAsnGlnAlaMetThrGlnLeuLysGlnGlyIleAlaAspLysAsp 3562
3649 .....GACCGCCGCAA 3659
3563 GlnThrLysAlaAsnGlyAsnPheValAsnAlaAspThrAspLysGlnAs 3579
3660 CGCCGTTTGGACAAGCGGCA.....TCCGGGACA 3688
3579 nAla.TyrAsnAsnAlaValAlaHisAlaGluGlnIleIleSerGlyThr 3595
3689 CCAAACACTACC..... 3700
3596 ProAsnAlaAsnValAspProGlnGlnValAlaGlnAlaLeuGlnGlnVa 3612
3701 .....GTTCGCA 3707
3612 lAsnGlnAlaLysGlyAspLeuAsnGlyAsnHisAsnLeuGlnValAlaL 3629
3708 AGATTTCCGCGCCTACCGCC.....AACAAACCG 3736
3629 ysAspAsnAlaAsnThrAlaIleAspGlnLeuProAsnLeuAsnGlnPro 3645
3737 ACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGC 3786
3646 .....GlnLysThrAlaLeuLysAspGlnValSe 3655
3787 ATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACGGCATCGG 3836
3655 rHisAlaGluLeuValThrGlyValAsnAlaIleLysGlnAsnAlaAspA 3672
3837 CAACTCGGCACGGCTTGCCCACG..... 3859
3672 laLeuAsnAsnAlaMetGlyThrLeuLysGlnGlnIleGlnAlaAsnSer 3688
3860 ...GTGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCG..... 3895
3689 GlnValProGlnSerValAspPheThrGlnAlaAspGlnAspLysGlnGl 3705
3896 .....ACATCGGCATCAGCGCGGGCGCGGGTTTTA..... 3925
3705 nAlaTyrAsnAsnAlaAlaAsnGlnAlaGlnGlnIleAlaAsnGlyIleP 3722
3926 .....GTAGCGGCAGCCTTTCAGACGGCA 3949
3722 roThrProValLeuThrProAspThrValThrGlnAlaValThrThrMet 3738
3950 TCAGAGGCAAAATCCGCC..... 3967
3739 AsnGlnAlaLysAspAlaLeuAsnGlyAspGluLysLeuAlaGlnAlaLy 3755
3968 .....GCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCGCGCAGG 4010
3755 sGlnGluAlaLeuAlaAsnLeuAspThrLeuArg...AspLeuAsnGlnP 3771
4011 TTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTC. 4059
3771 roGlnArgAspAlaLeuArgAsnGlnIleAsnGlnAlaGlnAlaLeuAla 3787
4060 ...GTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGCCACCCC 4106

```
3788 ThrValGluGlnThrLysGlnAsnAlaGlnAsnValAsnThrAlaMetSe 3804

4107 GGGCCTTGCATTCAACCGCTACCGCGCGGGCATT.............. 4140
      :::|||               :::   ||||||
3804 rAsnLeu.............LysGlnGlyIleAlaAsnLysAspThrV 3816

4141 ..AAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCT 4188
       ||||||:::   ::::::      |||:::      :::   |||
3816 alLysAlaSerGluAsnTyrHisAspAlaAspAlaAspLysGlnThr... 3831

4189 TATTTGAGCCTGTCCTATACCGATGCC.........GCTTCCGGCAAAGT 4229
             :::|||||||:::|||         |||:::|||   ::
3832 ...........AlaTyrThrAsnAlaValSerGlnAlaGluGlyIleIl 3844

4230 CCGAACGCGCGTCAATACCGCCGTA...........TTGGCGCAGGATT 4267
     ::::      :::|||   ::::::           :::::::::   :
3844 eAsnGlnThrThrAsnProThrLeuAsnProAspGluIleThrArgAlaL 3861

4268 TCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAA 4314
     ::   ::::::   :::|||:::   |||:::|||:::|||   |||
3861 euThrGlnValThrAspAlaLysAsnGlyLeuAsnGlyGluAlaLys 3876
```